Um estudo comparativo entre abordagens Bayesianas à testes de hipóteses

Brian Alvarez Ribeiro de Melo

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. Luís Gustavo Esteves

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

São Paulo, março de 2013

Um estudo comparativo entre abordagens Bayesianas à testes de hipóteses

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 04/03/2013. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Luís Gustavo Esteves (orientador) IME-USP
- Prof. Dr. Adriano Polpo de Campos DEs-UFSCar
- Prof^a. Dr^a. Verónica Andrea González López Unicamp

Agradecimentos

Em primeiro lugar, agradeço a Deus por me dar a coragem e a força necessárias ao longo de todos esses anos.

Agradeço a minha família, especialmente aos meus pais, Mônica e Ricardo, por todo o apoio e incentivo recebidos.

Agradeço ao meu orientador e amigo, Professor Luís Gustavo, pelos ensinamentos, disposição e paciência durante o desenvolvimento deste trabalho.

Resumo

MELO, B. A. R. Um estudo comparativo entre abordagens Bayesianas à testes de hipóteses. 2013. 80 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.

Neste trabalho, consideramos uma população finita composta por N elementos, sendo que para cada unidade está associado um número (ou vetor) de tal forma que temos para a população o vetor de valores $X = (X_1, \ldots, X_N)$, onde X_i denota a característica de interesse do i-ésimo indivíduo da população, que suporemos desconhecida. Aqui assumimos que a distribuição do vetor X é permutável e que existe disponível uma amostra composta por n < N elementos. Os objetivos são a construção de testes de hipóteses para os parâmetros operacionais, através das distribuições a posteriori obtidas sob a abordagem preditivista para populações finitas e a comparação com os resultados obtidos a partir dos modelos Bayesianos de superpopulação. Nas análises consideramos os modelos Bernoulli, Poisson, Uniforme Discreto e Multinomial. A partir dos resultados obtidos, conseguimos ilustrar situações nas quais as abordagens produzem resultados diferentes, como prioris influenciam os resultados e quando o modelo de populações finitas apresenta melhores resultados que o modelo de superpopulação.

Palavras-chave: inferência bayesiana, preditivismo, teste de hipóteses, populações finitas.

Abstract

MELO, B. A. R. A comparative study of Bayesian approaches to hypothesis testing. 2013. 80 f. MSc thesis - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.

We consider a finite population consisting of N units and to each unit there is a number (or vector) associated such that we have for the population the vector of values $X = (X_1, \ldots, X_N)$, where X_i denotes the characteristic of interest of the i-th individual in the population, which we will suppose unknown. Here we assume that the distribution of the vector X is exchangeable and that there is available a sample of size n < N from this population. The goals are to derive tests of hipotheses for the operational parameters through the corresponding posterior distributions obtained under the predictivistic approach for finite populations and to compare them with the results obtained from the usual Bayesian procedures of superpopulation models. In the analysis, the following models are considered: Bernoulli, Poisson, Discrete Uniform and Multinomial. From the results, we can identify situations in which the approaches yield different results, how priors influence the results of hipothesis testing and when the finite population model performs better than the superpopulation one.

Keywords: bayesian inference, predictivism, hipothesis testing, finite population.

Sumário

1	Introdução					
	1.1	Abordagem Preditivista	1			
	1.2	Objetivos	3			
	1.3	Organização do Trabalho	3			
2	Permutabilidade e Teorema da Representação					
	2.1	Permutabilidade	5			
	2.2	Teorema da Representação de de Finetti	6			
3	Testes de Hipóteses Bayesianos 9					
	3.1	Testes de Bayes	9			
	3.2	FBST	11			
	3.3	FBST para distribuições discretas	12			
4	Distribuições Preditivas					
	4.1	Distribuição Bernoulli	16			
		4.1.1 Performance dos testes	25			
	4.2	Distribuição Poisson	26			
	4.3	Distribuição Uniforme em $\mathbb N$	35			
	4.4	Distribuição Multinomial	42			
		4.4.1 Comparação entre duas proporções	45			
		4.4.2 Testando proporções simultaneamente	51			
5	Considerações Finais					
	5.1	Perspectivas Futuras	57			
A	Bernoulli					
в	3 Poisson					
\mathbf{C}	C Multinomial					
Re	e ferê :	ncias Bibliográficas	69			

x SUMÁRIO

Capítulo 1

Introdução

1.1 Abordagem Preditivista

A concepção usual de modelos estatísticos que dependem de parâmetros arbitrários, embora muito utilizada na prática, é muito discutida por vários autores. Por exemplo, Wechsler (1993) diz que, para um estatístico frequentista, a idéia de probabilidades fixas, mas desconhecidas, que existem de alguma forma, leva à suposição da existência de parâmetros.

De Finetti dizia que só podemos atribuir probabilidades à quantidades observáveis e que modelos paramétricos fazem apenas a ligação da experiência passada com a incerteza sobre possíveis novas observações (Fortini e Petrone, 2012). Schervish (1995) diz que, sob a perspectiva clássica, em um processo de variáveis 0-1, um estatístico inventa um parâmetro, θ , o qual assume-se ser um valor fixo e desconhecido tal que $P(X_i = 1) = \theta$. Sob a perspectiva Bayesiana, o estatístico construiria uma distribuição de probabilidades μ para θ e diria que:

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)} d\mu(\theta),$$
(1.1)

 $\forall n \in \{1, 2, ...\}, \forall x_1, ..., x_n \in \{0, 1\}$. Muito se discute sobre o que seria esta quantidade θ e sobre sua existência. Através do Teorema de Representação para variáveis 0-1 de de Finetti, obtemos outra interpretação, diferente da frequentista, para o parâmetro θ . Esta quantidade é o limite quase certo da sequência de frequências relativas construída a partir da sequência infinita, $(X_n)_{n\geq 1}$, de variáveis 0-1.

Por outro lado, considerando uma abordagem de populações finitas, utilizamos parâmetros operacionais, os quais são definidos como funções dos valores da população. Por exemplo, considere a seguinte situação: uma urna contém N bolas, sendo que em cada bola está marcado ou o valor 0 ou 1. São conhecidos os valores de n < N bolas desta urna. O interesse é descobrir a quantidade (proporção) de bolas marcadas com o número 1. O parâmetro operacional será então definido como $T(X) = \sum_{i=1}^{N} X_i (T'(X)) = \sum_{i=1}^{N} X_i / N)$, o qual é uma quantidade observável, pois, se retirarmos todas as bolas da urna (o que pode ser feito em um número finito de passos), saberemos o verdadeiro valor do parâmetro operacional T(X).

De forma mais geral, em diversas situações temos uma população finita e fechada (no sentido de que não existem quantidades não observáveis (Iglesias *et al.*, 2009)), $P = \{1, 2, ..., N\}$, e, para cada unidade populacional, está associado um número real (ou vetor) desconhecido, X_i . Desta forma, temos para a população o vetor de valores (vetores) desconhecidos $X = (X_1, X_2, ..., X_N)$, onde X_i denota a característica (vetor de características) de interesse do i-ésimo indivíduo da população (usualmente $X_i \in \mathbb{R}^k, k \ge 1$), i = 1, ..., N. Assim, a única aleatoriedade existente está nos valores desconhecidos de X e podemos definir, a partir de X, os parâmetros operacionais. Nesse caso, considerando disponível uma amostra de tamanho n < N, o procedimento para realizarmos inferência sobre os parâmetros operacionais é chamado *preditivista* ou *Bayes operacional* e foi introduzido por de Finetti (Iglesias *et al.*, 2009). Nesta abordagem, probabilidades são atribuídas apenas ao vetor X.

A vantagem de trabalharmos com a abordagem preditivista e, portanto, com parâmetros operacionais, é que tais quantidades são consideradas observáveis, ao invés de quantidades abstratas como limites de frequências relativas e, assim, não há necessidade de definir um espaço paramétrico (Mendel, 1994).

Embora, em algumas situações, o modelo de populações finitas possa ser estendido para um modelo de superpopulação, isso nem sempre é possível (Iglesias, 1993) e, mesmo em situações nas quais o modelo preditivista é estendível, existem diferenças entre as abordagens. Iglesias (1993) discute condições de estendibilidade para alguns modelos preditivistas.

Neste trabalho, consideramos o total populacional, dado por $T(X) = \sum_{i=1}^{N} X_i$ e o máximo populacional, $X_{(N)} = max\{X_1, \ldots, X_N\}$, como os parâmetros operacionais de interesse nos casos univariados. No caso multivariado de dimensão M, o parâmetro operacional é o vetor de totais populacionais $T(X) = \sum_{i=1}^{N} X_i = (T_1, \ldots, T_M)$, no qual T_i denota o total populacional da i-ésima característica de interesse.

1.2 Objetivos

Considerando o contexto descrito acima e a abordagem preditivista para populações finitas, neste trabalho:

- estudamos a abordagem preditivista para populações finitas em alguns modelos usuais e a construção das distribuições a priori e a posteriori para os parâmetros operacionais à partir da suposição de permutabilidade e condições adicionais de simetria;
- estudamos a construção de testes de hipóteses Bayesianos, aplicando-os aos modelos preditivistas de populações finitas e comparando-os aos correspondentes procedimentos de superpopulação. Consideramos o teste de Bayes (baseado nas probabilidades a posteriori das hipóteses) no caso de hipóteses compostas e não-precisas e o Teste de Significância Genuinamente Bayesiano (*Full Bayesian Significance Test* FBST) para hipóteses precisas.

1.3 Organização do Trabalho

No Capítulo 2, apresentamos a definição de permutabilidade e o Teorema da Representação de de Finetti e discutimos suas interpretações e aplicações. As definições dos testes bayesianos mais usuais para hipóteses precisas e compostas (unilaterais) são apresentadas no Capítulo 3. No Capítulo 4, apresentamos as distribuições a posteriori obtidas sob a abordagem preditivista para populações finitas. Os testes de hipóteses bayesianos são aplicados aos parâmetros operacionais e aos parâmetros dos modelos de superpopulação gerados a partir das mesmas suposições de simetria. Por fim, são feitas algumas considerações finais sobre os resultados obtidos e discutidas algumas propostas futuras de pesquisa.

4 INTRODUÇÃO

Capítulo 2

Permutabilidade e Teorema da Representação

2.1 Permutabilidade

Ao observarmos uma sequência de variáveis aleatórias X_1, \ldots, X_N , podemos assumir, em certas situações (como no exemplo do lançamento de percevejos abaixo), que a mudança de ordem dos valores das variáveis não altera a incerteza a respeito do resultado da sequência, ou seja, os rótulos das variáveis são não informativos (Bonassi, 2009). Esta é a idéia de permutabilidade finita, cuja definição formal é apresentada a seguir.

Definição 2.1 (Pemutabilidade Finita). Uma sequência finita de variáveis aleatórias X_1, \ldots, X_N assumindo valores em \mathfrak{X} é dita permutável se a distribuição conjunta f(.) desta sequência satisfaz

$$f(x_1,\ldots,x_N) = f(x_{\pi(1)},\ldots,x_{\pi(N)}), \quad \forall x_i \in \mathfrak{X},$$
(2.1)

para todas as permutações π definidas no conjunto $\{1, \ldots, N\}$.

A definição de permutabilidade finita nos leva a pensar na extensão para uma sequência infinita de variáveis aleatórias. Desta forma, temos a seguinte definição:

Definição 2.2 (Pemutabilidade Infinita). Uma sequência infinita de variáveis aleatórias é permutável se toda subsequência finita dela for permutável.

Esta definição é importante para entendermos o Teorema da Representação de de Finetti, que será apresentado na próxima seção.

Lindley e Phillips (1976) discutem um exemplo sobre o lançamento de um percevejo (pequeno prego, de cabeça redonda, usado para fixar papel). Neste experimento, o resultado é descrito através de variáveis binárias, X_1, \ldots, X_n , onde $X_i = 1$ se o percevejo cai com a ponta para cima no i-ésimo lançamento e $X_i = 0$, caso contrário, $i = 1, \ldots, n$. Neste cenário, parece natural considerarmos que a informação atribuída ao resultado de um lançamento seja encarada como equivalente à informação proveniente do resultado de qualquer outro lançamento. Do mesmo modo, entenderíamos a informação obtida através de dois lançamentos específicos do percevejo como equivalente à informação dos resultados de quaisquer outros dois lançamentos: o mesmo ocorre ao considerarmos três lançamentos, quatro lançamentos, e assim por diante. Essa simetria nos resultados dos lançamentos pode ser expressa pela suposição de permutabilidade da sequência de lançamentos. Iglesias (1993) discute diversas propriedades de variáveis permutáveis e mostra que todas distribuições marginais de subcoleções de variáveis aleatórias permutáveis também são permutáveis.

2.2 Teorema da Representação de de Finetti

Os Teoremas de Representação, primeiramente desenvolvidos por de Finetti, estão entre os mais importantes para a inferência estatística, pois eles fornecem a interpretação dos parâmetros sob o ponto de vista bayesiano. A seguir, enunciamos o caso mais simples destes teoremas, o Teorema da Representação para variáveis 0-1 de de Finetti.

Teorema 2.1 (Teorema da Representação para variáveis 0-1). Seja $(X_n)_{n\geq 1}$ uma sequência infinita permutável de variáveis aleatórias assumindo valores em $\{0,1\}$. A distribuição de qualquer subsequência de comprimento finito $n, n \geq 1$, pode ser expressa como

$$f(x_1, \dots, x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} d\mu(\theta) , \qquad (2.2)$$

para alguma medida de probabilidade μ concentrada no intervalo $[0,1], x_1, \ldots, x_n \in \{0,1\}.$

Além disso,

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{qc} \theta ,$$

sendo que θ é distribuído segundo a medida μ .

A demonstração deste teorema foi apresentada por diversos autores e pode ser encontrada, entre outros, em Heath e Sudderth (1976) e Schervish (1995). Os teoremas de representação caracterizam as distribuições conjuntas de variáveis aleatórias permutáveis. A idéia central destes teoremas é que se $(X_n)_{n\geq 1}$ é uma sequência infinita de variáveis aleatórias permutáveis, então existe uma medida \mathbb{P} tal que, condicionado em \mathbb{P} , as variáveis $(X_n)_{n\geq 1}$ são condicionalmente independentes e identicamente distribuídas segundo \mathbb{P} .

Esta representação mostra que a medida de probabilidade da sequência permutável de variáveis $0 - 1 \ (X_n)_{n\geq 1}$ pode ser expressa como uma mistura de medidas de probabilidade associadas a sequências de variáveis aleatórias Bernoulli independentes e identicamente distribuídas, ponderadas pela distribuição a priori, μ .

Considerando o exemplo dos percevejos discutido na seção anterior e assumindo a permutabilidade para uma possível sequência infinita de lançamentos, então o Teorema 2.1 dá um significado ao parâmetro do modelo Bernoulli(θ), como sendo uma variável aleatória, a saber, o limite (infinito) das frequências relativas de vezes que o percevejo cai com a ponta para cima e não uma quantidade fixa (não aleatória), como na perspectiva clássica. Além disso, a hipótese de permutabilidade fica clara na representação em (2.2), uma vez que todas as permutações de (x_1, \ldots, x_n) produzem o mesmo valor de $\sum_{i=1}^n x_i$. Desse modo, a expressão em (2.2) exibe uma representação (sob forma integral) da avaliação subjetiva da permutabilidade em um processo de variáveis 0 – 1, segundo a visão preditivista. Diversos Teoremas de Representação foram desenvolvidos para justificar outros modelos paramétricos segundo a visão preditivista. Resultados desta forma são denominados "Formas Infinitas do Teorema de de Finetti"(Iglesias, 1993).

Por outro lado, estamos interessados no estudo de populações finitas, ou seja, situações em que o tamanho populacional N é conhecido. Assim, consideramos sequências finitas de variáveis aleatórias, situação na qual o Teorema de Representação não é necessariamente válido (Heath e Sudderth, 1976). Um exemplo disto é o caso de duas variáveis permutáveis, $X_1 \in X_2$, tais que $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 0, 5$. Para esses casos, em que não existe a "forma integral" (como em (2.2)) do processo, foram desenvolvidas as versões finitas do teorema de de Finetti (Schervish, 1995). Resultados envolvendo formas finitas do teorema de de Finetti foram apresentados por Diaconis (1977), Wechsler (1993), Iglesias (1993), entre outros.

No Capítulo 4, serão apresentadas as versões finitas do Teorema de de Finetti para algumas distribuições usuais. A partir destes resultados, encontramos as distribuições a posteriori dos parâmetros operacionais, as quais usaremos para construir os testes de hipóteses e realizar as comparações com os métodos em superpopulação.

Capítulo 3

Testes de Hipóteses Bayesianos

Testes de hipóteses podem ser vistos como procedimentos de decisão utilizados quando queremos fazer uma afirmação a respeito de uma quantidade desconhecida, usualmente um parâmetro, θ . Sob o ponto de vista bayesiano, os procedimentos de teste de hipóteses são construídos com base na *Teoria da Decisão*. Neste capítulo, apresentamos a construção e interpretação de dois testes de hipóteses bayesianos: o Teste de Bayes baseado nas probabilidades a posteriori das hipóteses e o Teste de Significância Genuinamente Bayesiano (*Full Bayesian Significance Test* - FBST).

Para testarmos hipóteses não precisas (Pereira *et al.*, 2008), escolhemos o Teste de Bayes por sua simplicidade, tanto em sua formulação quanto na interpretação dos resultados. No caso de hipóteses precisas, optamos pelo FBST por não introduzir uma probabilidade positiva para a hipótese nula nem tornar necessária a eliminação de parâmetros de perturbação (Pereira *et al.*, 2008). Outras alternativas também poderiam ser escolhidas, como testes baseados no fator de Bayes ou o teste de Jeffreys.

3.1 Testes de Bayes

Seja θ um parâmetro populacional e Θ o espaço paramétrico, conjunto de todos os possíveis valores para θ . Considere que existe uma distribuição de probabilidades a priori para θ e suponha disponível uma amostra (X_1, \ldots, X_n) , cuja distribuição, dado θ , é conhecida. Com base na incerteza a priori e na informação obtida pela amostra, queremos testar as seguintes hipóteses:

$$\begin{cases}
H_0: \theta \in \Theta_0 \\
H_1: \theta \in \Theta_1
\end{cases},$$
(3.1)

,

de tal forma que $\Theta = \Theta_0 \cup \Theta_1 \in \Theta_0 \cap \Theta_1 = \emptyset$.

Em um teste de hipóteses como descrito acima existem duas possíveis decisões: rejeiçãoou não rejeição da hipótese nula, denotadas por d_1 e d_0 , respectivamente. Consideramos, então, a seguinte função de perda para as possíveis decisões:

Tabela 3.1: Função de perda para teste de hipóteses

		$\theta\in\Theta_0$	$\theta\in\Theta_1$
(não rejeitar)	d_0	b_0	a_1
(rejeitar)	d_1	a_0	b_1

Nas funções de perda como a apresentada na Tabela 3.1, usualmente consideramos $b_0 = b_1 = 0$, ou seja, se decidimos por não rejeitar H_0 e, de fato, $\theta \in \Theta_0$, então não existe perda. O mesmo ocorre quando $\theta \in \Theta_1$ e decidimos pela rejeição de H_0 . Desta forma, a perda ao se tomar uma decisão correta é nula. Por outro lado, a perda associada ao Erro Tipo I, rejeição de H_0 quando esta é verdadeira, é a_0 , enquanto que a perda associada ao Erro Tipo II, não rejeição de H_0 quando deveríamos rejeitá-la, é igual a a_1 .

O Teste de Bayes é, nesse caso, o teste que para cada possível amostra minimiza o valor esperado a posteriori da perda da Tabela 3.1. Fossaluza (2008) mostra que o teste que rejeita H_0 se

$$P(\theta \in \Theta_0 | (x_1, \dots, x_n)) < \frac{a_1 - b_1}{(a_0 - b_0) + (a_1 - b_1)}$$

sendo $a_0 > b_0$ e $a_1 > b_1$, é o teste de Bayes. No caso particular de $b_0 = b_1 = 0$, rejeita-se H_0 quando a sua probabilidade a posteriori for menor que $\frac{a_1}{a_0+a_1}$.

A vantagem deste procedimento de teste é que basta calcularmos a probabilidade (a posteriori) da hipótese. Além disso, a interpretação do resultado do teste é bem simples: rejeita-se H_0 se sua probabilidade for "baixa"e aceita-se H_0 quando sua probabilidade for "alta", dependendo dos valores $a_0 e a_1$. Contudo, o teste de Bayes baseado nas probabilidades a posteriori das hipóteses não é o ideal para todas as situações.

3.2 FBST

O Teste de Bayes descrito anteriormente, embora possa ser aplicado em diversas situações, não é adequado aos casos em que o subconjunto do espaço paramétrico Θ_0 tem medida nula (DeGroot, 1970). Por exemplo, considere a situação na qual a distribuição a posteriori do parâmetro θ é contínua em \mathbb{R} para toda possível amostra x e queremos testar as seguintes hipóteses:

$$\begin{cases}
H_0: \theta = \theta_0 \\
H_1: \theta \neq \theta_0
\end{cases},$$
(3.2)

com $\theta_0 \in \mathbb{R}$, constante. Neste caso, temos que

$$P(\theta = \theta_0 | x) = 0 ,$$

independentemente da amostra x obtida e, assim, o teste de Bayes, se aplicado a esta situação, conduziria sempre à rejeição de H_0 , tornando-se, portanto, inadequado.

Pereira e Stern (1999) propuseram então o Full Bayesian Significance Test (FBST) para hipóteses precisas (subconjuntos de Θ que possuem dimensão inferior à dimensão do espaço paramétrico e, portanto, com medida nula). O FBST é baseado na evidência em favor da hipótese nula, calculada a partir da credibilidade da região tangente à H_0 , conforme a Definição 3.1 abaixo. Madruga *et al.* (2001) apresentam funções de perda que conferem "bayesianidade"ao FBST, isto é, tornam o FBST um legítimo procedimento bayesiano, sob o enfoque de Teoria da Decisão. A seguir, segue a definição formal do FBST:

Definição 3.1 (FBST). Seja Θ o espaço paramétrico, $\theta \in \Theta$ o parâmetro de interesse (θ é valor absolutamente contínuo), f(.) uma função densidade de probabilidade a priori sobre Θ , f(.|x) a função densidade a posteriori de θ dada uma amostra específica $x \in H_0 : \theta \in \Theta_0$ a hipótese nula a ser testada (considere que dim $(\Theta_0) < \dim(\Theta)$). A região tangente à hipótese nula é dada por:

$$T_x = \left\{ \theta \in \Theta : f(\theta|x) > \sup_{\Theta_0} f(\theta|x) \right\} .$$

A evidência em favor de H_0 é definida como:

$$ev(\Theta_0; x) = 1 - P(\theta \in T_x | x) = 1 - \int_{T_x} f(\theta | x) d\theta$$

O FBST é o teste que rejeita H_0 quando $ev(\Theta_0; x)$ for "pequena" (Pereira et al., 2008).

Após a apresentação do FBST por Pereira e Stern (1999), diversos trabalhos foram desenvolvidos no sentido de explorar propriedades e aplicações deste procedimento em testes de hipóteses, além de comparações com outros métodos para testes de hipóteses precisas, como o *fator de Bayes* e o teste de Jeffreys. Entre outros trabalhos, destacamos Madruga *et al.* (2001), Madruga e Pereira (2005), Pereira *et al.* (2008) e Izbicki (2010).

3.3 FBST para distribuições discretas

Na proposta de Pereira e Stern (1999), o FBST é apresentado como um teste a ser aplicado em situações nas quais a distribuição a posteriori do parâmetro de interesse é contínua e a hipótese nula é precisa. A seguir, apresentamos um exemplo como motivação para a extensão deste procedimento para os casos em que a distribuição do parâmetro de interesse é discreta.

Considere que possuímos uma urna com N = 100 bolas de cor branca ou preta. Definimos $X_i = 1$, se a i-ésima bola é da cor preta e $X_i = 0$, caso contrário, i = 1, ..., 100. O parâmetro (operacional) de interesse é $T = \sum_{i=1}^{N} X_i$, o total de bolas pretas na urna. Considere o experimento que consiste em retirar uma amostra de n = 20 bolas da urna. Suponha que encontramos um total de 10 bolas pretas. Assim, a previsão sobre T será baseada nesta amostra (suponha, sem perda de generalidade, que foram observados os valores X_1, \ldots, X_{20}). Formulamos, então, o modelo da seguinte maneira:

- Assumimos que X_1, \ldots, X_{100} são permutáveis e que a distribuição a priori para a quantidade de bolas pretas na urna é Uniforme Discreta no conjunto $\{0, 1, \ldots, 100\}$;
- Nesse caso, a probabilidade da amostra, condicionada ao total de bolas pretas T, é

dada por:

$$P\left(\sum_{i=1}^{20} X_i = j \middle| T = t\right) = \frac{\binom{t}{j}\binom{100-t}{20-j}}{\binom{100}{20}} \quad ;$$

• A distribuição a posteriori para T é :

$$P\left(T=t\left|\sum_{i=1}^{20}X_i=10\right)\propto \binom{t}{10}\binom{100-t}{10}\right).$$

Agora queremos saber se a quantidade de bolas pretas e brancas na urna é a mesma. Do ponto de vista estatístico, isso equivale a testar as seguintes hipóteses:

$$\begin{cases} H_0: T = 50 \\ H_1: T \neq 50 \end{cases}$$
(3.3)

Como a hipótese nula tem probabilidade a posteriori positiva, poderíamos pensar em utilizar o teste descrito na Seção 3.1. No entanto, a probabilidade a posteriori de H_0 , 0,04, pode ser considerada "baixa"para um procedimento deste tipo. Por outro lado, o ponto T = 50 é o mais verossímil e também com a maior probabilidade a posteriori (Figura 3.1).

Assim, faz sentido que não rejeitemos a hipótese nula, o que nos leva a pensar em um procedimento alternativo para esta situação. Baseado no FBST para modelos contínuos, podemos calcular a evidência em favor de H_0 da mesma forma (na verdade, podemos proceder desta forma em qualquer situação em que a distribuição de θ é dominada por uma medida σ -finita (Schervish, 1995)) apenas realizando as devidas adaptações para funções de probabilidade discretas. Assim, considerando hipóteses da forma em (3.2) e que a distribuição a posteriori do parâmetro θ é discreta, temos que a evidência em favor de $H_0: \theta = \theta_0$ é dada por:

$$ev(\{\theta_0\}; x) = 1 - \sum_{\{k: P(\theta=k|x) > P(\theta=\theta_0|x)\}} P(\theta=k|x)$$
 (3.4)

e rejeitamos a hipótese nula sempre que o valor de $ev(\{\theta_0\}; x)$ for "pequeno".

Considerando este procedimento para o exemplo, obtemos $ev({50}; 10) = 1$, pois o ponto T = 50 é a moda da posteriori e, portanto, não existem pontos no espaço paramétrico com



Figura 3.1: Distribuição a posteriori de T

maior probabilidade, fazendo com que a região tangente à hipótese nula seja o conjunto vazio. Considerando um ponto de corte igual a 0,1, decidimos por não rejeitar H_0 , concluindo que não temos evidência suficiente para afirmar que as quantidades de bolas brancas e pretas na urna são diferentes.

Note que, assim como na formulação original do FBST, para a construção da região tangente à hipótese nula consideramos apenas os pontos com probabilidade estritamente maior que a probabilidade da moda restrita à H_0 . Dessa forma, a evidência em favor da hipótese nula, quando esta hipótese contém a moda da posteriori, será igual a 1, assim como no caso contínuo.

Assim, no Capítulo 4, testamos, sob o modelo de superpopulação, hipóteses precisas através do FBST em sua formulação original e hipóteses não precisas através do teste apresentado na Seção 3.1. Sob o modelo de populações finitas, adotamos as correspondentes versões discretas destes testes. Isso viabiliza uma melhor comparação entre os resultados dos testes sob os dois modelos.

Capítulo 4

Distribuições Preditivas

Como discutido nos capítulos anteriores, a abordagem preditivista foi introduzida por de Finetti e é vista como uma forma "mais pura"da abordagem Bayesiana, na qual o interesse é realizar inferência apenas sobre quantidades observáveis. Neste capítulo, são construídos testes de hipóteses para os parâmetros de interesse sob a perspectiva Bayesiana usual de superpopulações e considerando a abordagem preditivista para populações finitas.

Modelos de populações finitas são usualmente utilizados quando conhecemos a quantidade de elementos presentes na população em estudo, ou seja, N é conhecido. Quando não sabemos o valor de N, utilizamos os modelos de superpopulação. Uma aplicação de modelos de populações finitas consiste nas sucessivas avaliações da distribuição de votos dos candidatos de um pleito eleitoral a partir das contagens parciais de votos, obtidas ao longo do processo de apuração. Nesta situação, a quantidade de eleitores, N, é conhecida, e as diversas contagens parciais de votos podem ser interpretadas como amostras, de tamanhos pequenos (no início da apuração) e tamanhos grandes em relação ao valor de N, quando a apuração já atinge boa parte do total de votos. Nesses últimos casos, quando o tamanho amostral é grande relativamente ao populacional, o modelo de superpopulação pode conduzir a erros grandes se comparados aos modelos de população finita.

Em geral, para a construção das distribuições a posteriori sob a abordagem preditivista, supomos a existência de uma população de interesse, $P = \{1, ..., N\}$, com N conhecido, a qual está associado o vetor permutável de valores desconhecidos $X = (X_1, ..., X_N), X_i \in \mathbb{R}^k$, $k \geq 1$. Além disso, existe disponível uma amostra de n < N elementos da população e, sem perda de generalidade, consideramos que esta amostra é composta pelos n primeiros elementos da população, (X_1, \ldots, X_n) .

Nesse trabalho, são considerados testes de hipóteses, tanto sob a abordagem preditivista para populações finitas como sob o enfoque de superpopulação, para os seguintes modelos usuais de superpopulação: *Bernoulli, Poisson, Uniforme Discreto e Multinomial*.

4.1 Distribuição Bernoulli

A inferência sobre uma proporção é um problema muito usual em várias áreas do conhecimento, de tal maneira que podemos imaginá-la em diversas aplicações, tais como descobrir a quantidade (proporção) de bolas pretas em uma urna que contenha um total de N bolas ou a porcentagem de eleitores de uma localidade que votarão em determinado candidato em uma eleição.

Nestas situações, podemos considerar $X = (X_1, \ldots, X_N)$ tal que $X_i \in \{0, 1\}$, $i = 1, \ldots, N$. Desta forma, o interesse está em fazer inferência sobre a proporção de 1's (sucessos) na população. Por simplicidade, consideramos o parâmetro operacional $\theta = T(X) = \sum_{i=1}^{N} X_i$, de modo que a proporção de sucessos seja $\pi = \frac{\theta}{N}$. Portanto, esses parâmetros são equivalentes, no sentido que $P(\theta = j) = P(\pi = \frac{j}{N}), \forall j = 0, 1, \ldots N$. Dispondo da realização da amostra $(X_1, \ldots, X_n), n < N$, temos o seguinte resultado acerca da distribuição a posteriori de θ :

Resultado 4.1. Suponha que o vetor $X = (X_1, ..., X_N)$ assumindo valores em $\{0, 1\}^N$ é permutável e defina $P(\theta = j) = p_j$, j = 0, 1, ..., N, como a incerteza a priori a respeito do parâmetro θ . A distribuição a posteriori de θ , dado a amostra $(x_1, ..., x_n)$, é dada por:

$$P\left(\theta = t \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \frac{\binom{N-n}{t-\sum_{i=1}^n x_i}}{\binom{N}{t}} p_t, \tag{4.1}$$

se $t \ge \sum_{i=1}^{n} x_i$. Além disso, $P(\theta = t \mid \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i) = 0$, para $t < \sum_{i=1}^{n} x_i$.

Demontração: Da hipótese de permutabilidade, temos (Iglesias, 1993):

$$P(X_1 = x_1, \dots, X_N = x_N | \theta = t) = \frac{1}{\binom{N}{t}} I_{\{t\}}(\sum_{i=1}^N x_i)$$

Seja $\boldsymbol{x} = (x_1, \dots, x_n) \operatorname{com} x_i \in \{0, 1\}, i = 1, \dots, n, e \operatorname{considere o conjunto} S = \left\{ (v_{n+1}, \dots, v_N) \in \{0, 1\}^{N-n} : \sum_{i=n+1}^{N} v_i = t - \sum_{i=1}^{n} x_i \right\}$. A probabilidade da amostra \boldsymbol{x} , dado $\theta = t$, é:

$$P(X_{1} = x_{1}, \dots, X_{n} = x_{n} | \theta = t) =$$

$$= \sum_{S} P(X_{1} = x_{1}, \dots, X_{n} = x_{n}, X_{n+1} = v_{n+1}, \dots, X_{N} = v_{N} | \theta = t)$$

$$= \sum_{S} P(X_{1} = 1, \dots, X_{t} = 1, X_{t+1} = 0, \dots, X_{N} = 0 | \theta = t)$$

$$= \sum_{S} \frac{1}{\binom{N}{t}} = \frac{\binom{N-n}{t-\sum_{i=1}^{n} x_{i}}}{\binom{N}{t}}.$$

Finalmente, a distribuição a posteriori para θ é dada por:

$$P\left(\theta = t \mid X_1 = x_1, \dots, X_n = x_n\right) \propto$$
$$\propto P\left(X_1 = x_1, \dots, X_n = x_n \mid \theta = t\right) p_t$$
$$= \frac{\binom{N-n}{t-\sum_{i=1}^n x_i}}{\binom{N}{t}} p_t I_{\left\{\sum_{i=1}^n x_i, \sum_{i=1}^n x_i+1, \dots, N\right\}} (t)$$

A distribuição em (4.1) é a distribuição a posteriori para θ sob a abordagem preditivista para populações finitas. Observe que, no caso de variáveis binárias 0-1, a hipótese de permutabilidade é suficiente para a especificação completa do modelo (Iglesias (1993) e Iglesias *et al.* (2009)). Para outros modelos mais complexos, torna-se necessária a consideração de suposições "mais fortes", como, por exemplo, a suposição de multinomialidade no caso da Poisson (Wechsler, 1993), que será apresentada na próxima seção. Além disso, a distribuição a posteriori de θ em (4.1) depende da amostra apenas através do total amostral $\sum_{i=1}^{n} x_i$ e pode-se mostrar facilmente que coincide com a distribuição de θ dado $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i$.

Por outro lado, considerando um modelo de superpopulação, podemos atribuir à proporção de interesse uma distribuição Beta. Neste caso, assumimos que, a priori, $\pi \sim Beta(\alpha, \beta)$, $\alpha, \beta > 0$. Supomos ainda, que a sequência de variáveis aleatórias $(X_n)_{n\geq 1}$, dado π , são condicionalmente independentes e identicamente distribuídos (ciid) segundo a distribuição Bernoulli (π) , isto é, $(X_n)_{n\geq 1}|\pi \stackrel{ciid}{\sim} Ber(\pi)$. Desta forma, $\pi|(x_1,\ldots,x_n) \sim Beta(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$. Aqui, π representa a proporção de 1's no limite, isto é, quando o tamanho da população vai para infinito $(N \to \infty)$.

Note que, sob o enfoque de superpopulação, a distribuição de π é contínua, permitindo que π assuma qualquer valor no intervalo (0, 1), enquanto que, sob a abordagem preditivista, a proporção de interesse assume um número finito de valores, de modo que existem N + 1possíveis valores a priori e $N + 1 - \sum_{i=1}^{n} x_i$ após a amostra ser observada.

Nosso interesse agora está na construção de testes para as seguintes hipóteses:

$$\begin{cases}
H_0: \pi \le \pi_0 \\
H_1: \pi > \pi_0
\end{cases}$$
(4.2)

е

$$\begin{cases} H'_{0} : \pi = \pi_{0} \\ H'_{1} : \pi \neq \pi_{0} \end{cases},$$
(4.3)

para $\pi_0 \in (0, 1)$, através das duas abordagens estudadas e a comparação dos resultados desses testes.

Para testar as hipóteses em (4.2) e (4.3), utilizaremos o teste de Bayes baseado nas probabilidades a posteriori das hipóteses e o FBST, respectivamente. Para testarmos as hipóteses em (4.3) através da posteriori Beta no modelo de superpopulação, utilizaremos o FBST na sua formulação original. No entanto, como mencionado anteriormente, a distribuição a posteriori para a quantidade de sucessos sob a abordagem preditivista para populações finitas é discreta e, neste caso, utilizaremos a extensão do FBST para funções de probabilidade discretas, descrita na Seção 3.3. Note que o interesse está apenas na comparação, ainda que de maneira preliminar, dos valores das evidências (ou probabilidades) em favor da hipótese nula obtidos pelas duas abordagens, sendo que não serão discutidos critérios acerca de pontos de corte para decidirmos quando devemos rejeitar ou não alguma hipótese.

Nos exemplos a seguir, consideramos o valor $\pi_0 = 0, 5$. Desta forma, considerando as hipóteses unilaterais, a probabilidade de $H_0: \pi \leq 0, 5$, segundo o modelo de superpopulação é dada por:

$$P\left(H_0\left|\sum_{i=1}^n X_i=s\right)\right) = \int_0^{0.5} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s)\Gamma(\beta+n-s)} \pi^{\alpha+s-1} (1-\pi)^{\beta+n-s-1} d\pi$$

Contudo, sob a abordagem preditivista, a hipótese nula equivalente é $H_0: \theta \leq N/2$ e sua probabilidade a posteriori é dada por:

$$P\left(H_0\left|\sum_{i=1}^n X_i = s\right) = \frac{1}{c}\sum_{t=s}^{\lfloor\frac{N}{2}\rfloor} \frac{\binom{N-n}{t-s}}{\binom{N}{t}} p_t ,$$

sendo c a constante normalizadora, $c = \sum_{t=s}^{N} P(X_1 = x_1, \dots, X_n = x_n | \theta = t) p_t$, e $\lfloor b \rfloor$ o maior inteiro menor ou igual a b.

No caso da hipótese precisa H'_0 : $\pi = \pi_0$, sob o enfoque de superpopulação, a região tangente à hipótese nula e a evidência em favor de H_0 , obtidas segundo o FBST, são dadas por

$$T_{x} = \left\{ \pi \in (0,1) : f(\pi|s) > f(\pi_{0}|s) \right\}$$

e
 $ev(\{\pi_{0}\}; \boldsymbol{x}) = 1 - \int_{T_{x}} \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + s)\Gamma(\beta + n - s)} \pi^{\alpha + s - 1} (1 - \pi)^{\beta + n - s - 1} d\pi ,$

sendo f(.|s) a função densidade da posteriori $Beta(\alpha + s, \beta + n - s)$. Por outro lado, consi-

derando $\theta_0 = N\pi_0 \in \mathbb{N}$, sob o enfoque preditivista para populações finitas, temos

$$T_{x} = \left\{ k \in \{s, \dots, N\} : P\left(\theta = k \left| \sum_{i=1}^{n} X_{i} = s \right) > P\left(\theta = \theta_{0} \left| \sum_{i=1}^{n} X_{i} = s \right) \right\} \right\}$$

e
$$ev\left(\{\theta_{0}\}; \boldsymbol{x}\right) = 1 - \sum_{T_{x}} P\left(\theta = k \left| \sum_{i=1}^{n} X_{i} = s \right) \right].$$

Para a comparação dos resultados dos testes de hipóteses utilizando a abordagem preditivista para populações finitas e o modelo de superpopulação, o primeiro passo foi atribuir os valores de N e n. Consideramos, então, N = 100 e 1000, e tamanhos amostrais iguais a 20% e 80% do tamanho da população. Para a escolha das distribuições a priori para o parêmetro de interesse, consideramos, inicialmente, uniformidade no espaço paramétrico, isto é, assumimos todos os possíveis resultados como 'equiprováveis' e modelamos a incerteza a priori através das seguintes distribuições:

- $\theta \sim Uniforme\{0, 1, \dots, N\}$ abordagem preditivista;
- $\pi \sim Beta(1,1)$ modelo de superpopulação;

Finalmente, para cada possível valor $s \in \{0, 1, ..., n\}$ da estatística suficiente $\sum_{i=1}^{n} X_i$, calculamos a probabilidade da hipótese nula H_0 (ou a evidência em favor de H'_0).

Para ilustrarmos as diferenças entre as abordagens, são apresentadas as Figuras 4.1 e 4.2, as quais representam as probabilidades e evidências em favor das hipóteses nulas em função do total amostral, considerando uma população de tamanho N = 100, $\pi_0 = 0,5$ e amostras de tamanho 20 e 80. Resultados obtidos a partir de outros valores de $N e \pi_0$ são apresentadas no Apêndice A. Quando variamos o valor de π_0 , o comportamento das curvas apresentadas continua o mesmo, sendo que elas são apenas deslocadas em relação aos valores do total amostral.

Na Figura 4.1, os pontos em cinza representam as probabilidades da hipótese nula em função do total amostral obtidas através da distribuição em (4.1) e os pontos em preto as probabilidades obtidas pela posteriori Beta. Observamos que, quando a amostra é pequena em relação à população, os métodos produzem resultados próximos. Por outro lado, ao aumentarmos o tamanho amostral, nota-se que as diferenças entre as abordagens aumentam, sendo que nos casos em que o total amostral é igual a 37 ou 38, a diferença entre as probabilidades obtidas através dos dois métodos é superior a 20%.



Figura 4.1: Valores da probabilidade de $H_0: \pi \leq 0, 5$ ($\theta \leq \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 100.



Figura 4.2: Valores da evidência em favor de H'_0 : $\pi = 0,5$ ($\theta = \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 100.

Um comportamento similar ocorre para as evidências das hipóteses simples (precisas), apresentadas na Figura 4.2: os métodos produzem resultados próximos quando a amostra é pequena e as diferenças se destacam quando o tamanho amostral fica próximo ao populacional. Considerando a hipótese precisa $H'_0: \pi = 0, 5$, observamos, na Figura 4.2, que os valores da evidência obtidos sob a abordagem de populações finitas são sempre menores ou iguais às correspondentes evidências obtidas no modelo de superpopulação, ou seja, no modelo de populações finitas a tendência é que "rejeitemos mais" a hipótese nula.

Nas análises anteriores, foram utilizadas apenas prioris que não "privilegiam"nenhum ponto do espaço paramétrico, sendo a distribuição Uniforme discreta aplicada no caso preditivista e a Uniforme contínua (Beta(1,1)) para o parâmetro π da distribuição Bernoulli, no modelo de superpopulação. Para avaliar como prioris informativas afetam a forma da posteriori e, consequentemente, as probabilidades para os testes de hipóteses, consideramos prioris com diferentes médias. Para o caso de superpopulação, utilizamos a distribuição Beta(α, β), atribuindo diferentes valores para os parâmetros $\alpha \in \beta$. Para a abordagem de populações finitas, construímos a priori a partir da distribuição $Beta(\alpha, \beta)$ da seguinte forma: $\theta = \lfloor N\pi \rfloor$ (θ igual ao maior inteiro menor ou igual a $N\pi$), sendo $\pi \sim Beta(\alpha, \beta)$. Note que π é multiplicado pelo tamanho da população N, pois no caso preditivista o parâmetro é o total populacional, enquanto que no modelo de superpopulação consideramos a proporção. Desta forma, podemos comparar (ainda que em caráter exploratório) a influência de prioris informativas nos testes de hipóteses, pois, embora as distribuições a priori sejam diferentes, elas apresentam, num certo sentido, comportamentos semelhantes.

Na Figura 4.3, são apresentadas as probabilidades a posteriori da hipótese nula H_0 : $\pi \leq 0,5$ considerando diferentes distribuições a priori, uma população de tamanho N = 100e amostra com 20 elementos. Comparando os valores apresentados na Figura 4.3 com os calculados considerando prioris uniformes apresentado em 4.1(a), notamos que as diferenças das probabilidades a posteriori calculadas sob o modelo de superpopulação são sempre maiores ou iguais às diferenças das probabilidades calculadas sob o enfoque preditivista para populações finitas. Por exemplo, quando o total amostral é 8 e sob o enfoque de superpopulação, a probabilidade de H_0 diminui 0,45 quando consideramos uma distribuição beta com média 0,8, em relação ao modelo uniforme, enquanto que no modelo de populações finitas essa probabilidade diminui apenas 0,39, Figura 4.3(c). Um comportamento similar ocorre quando consideramos a distribuição *Beta*(2, 8) como priori para a proporção de sucessos. As diferenças são muito pequenas quando utilizamos a distribuição Beta(4, 4).

Aumentando o tamanho amostral para n = 80, observamos na Figura 4.4 que as diferenças entre as duas abordagens crescem. Comparando com os valores das probabilidades apresentados na Figura 4.1(b), constatamos novamente que as diferenças quando utilizamos o modelo de superpopulação são maiores que no modelo preditivista. Contudo, estas diferenças são menores do que quando utilizamos a amostra menor, com n = 20, chegando a no máximo 0,25 e 0,10 quando o total amostral é igual a 42, nos modelos de superpopulação e de população finita, respectivamente. Esses resultados indicam que o modelo de superpopulação é mais influenciado pelas distribuições a priori em relação ao modelo preditivista.

As evidências obtidas através dos testes para a hipótese H_0' , considerando diferentes distribuições a priori para π (θ), são apresentadas nas Figuras 4.5 e 4.6 para amostras de tamanho 20 e 80, respectivamente. Um comportamento similar ao dos resultados obtidos para hipóteses unilaterais ocorre: comparando as evidências obtidas utilizando prioris informativas com os resultados obtidos utilizando prioris uniformes, o modelo de superpopulação apresenta maiores diferenças. As diferenças nas evidências ao passarmos de uma priori uniforme para uma informativa são menores quando a amostra é maior. Em particular, essas diferenças são muito pequenas quando utilizamos a distribuição Beta(4, 4) como priori (caso em que as médias das prioris uniforme e informativa são iguais a 0,5).



Figura 4.3: Valores da probabilidade a posteriori de $H_0: \pi \leq 0, 5$ $(\theta \leq \frac{N}{2})$ em função do total amostral para uma população de tamanho N = 100 e amostra com n = 20 elementos.



Figura 4.4: Valores da probabilidade a posteriori de $H_0: \pi \leq 0, 5$ ($\theta \leq \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 100 e amostra com n = 80 elementos.



Figura 4.5: Valores da evidência em favor de $H'_0: \pi = 0, 5$ ($\theta = \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 100 e amostra com n = 20 elementos.



Figura 4.6: Valores da evidência em favor de $H'_0: \pi = 0, 5$ ($\theta = \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 100 e amostra com n = 80 elementos.

4.1.1 Performance dos testes

Nas análises conduzidas anteriormente, não foi estabelecido um verdadeiro valor para o total populacional, de tal forma que pudemos avaliar a probabilidade a posteriori da hipótese nula para todos os valores da estatística suficiente. A seguir, fixamos o valor do total populacional, o que claramente não implica em mudanças nos valores obtidos para as probabilidades a posteriori da hipótese nula, mas faz com que existam "menos amostras possíveis"e, consequentemente, "menos valores possíveis"para o total amostral. Por exemplo, no caso da Figura 4.7(a), consideramos N = 100, n = 80, $\pi_0 = 0.5$ ($\theta_0 = 50$) e $\sum_{i=1}^{N} X_i = 40$ e, desse modo, temos que $\sum_{i=1}^{n} X_i \in \{20, 21, \dots, 40\}$. Note que, caso considerássemos n = 20, o gráfico exibido na Figura 4.7(a) seria igual ao gráfico apresentado na Figura 4.1(a), pois o total amostral estaria no conjunto $\{0, 1, \dots, 20\}$. Na Figura 4.7(b), consideramos $\sum_{i=1}^{N} X_i =$ 60 (os demais valores não são alterados) e, desse modo, temos que $\sum_{i=1}^{n} X_i \in \{40, 41, \dots, 60\}$.

Na Figura 4.7(a), notamos que, para a maioria dos valores do total amostral, a probabilidade da hipótese nula $H_0: \pi \leq 0, 5$, sob a abordagem preditivista para populações finitas, é superior ou igual à probabilidade encontrada sob o enfoque de superpopulações, indicando que obtemos melhores resultados ao considerarmos o modelo de populações finitas, uma vez que o total populacional é igual a 40 e, portanto, não deveríamos rejeitar H_0 . Do mesmo modo, a Figura 4.7(b) mostra uma superioridade da abordagem preditivista, pois, neste caso, o ideal seria rejeitar H_0 , já que a proporção de sucessos na população é igual a 0,6 e as menores probabilidades da hipótese nula ocorrem na situação em que consideramos uma população finita.



Figura 4.7: Valores da probabilidade a posteriori de $H_0: \pi \leq 0, 5 \ (\theta \leq \frac{N}{2})$ em função do total amostral para N = 100 e n = 80.
4.2 Distribuição Poisson

O modelo Bernoulli, estudado na seção anterior, é o caso mais simples, pois as características de interesse assumem apenas os valores 0 e 1. Agora, consideramos um caso mais sofisticado, no qual as variáveis em estudo podem assumir qualquer valor no conjunto dos números naturais, $\mathbb{N} = \{0, 1, 2...\}$, isto é, $X = (X_1, \ldots, X_N) \in \mathbb{N}^N$ e o parâmetro de interesse é a média populacional, $\lambda = \frac{1}{N} \sum_{i=1}^{N} X_i$. No entanto, assim como no caso da distribuição Bernoulli, consideramos, sob a abordagem preditivista para populações finitas, o total populacional, $\Lambda = \sum_{i=1}^{N} X_i$, como o parâmetro operacional de interesse, uma vez que as distribuições destes parâmetros são equivalentes no sentido de que $P(\lambda = \frac{t}{N}) = P(\Lambda = t)$, $t \in \mathbb{N}$. Considerando disponível uma amostra (X_1, \ldots, X_n) , n < N, o resultado a seguir exibe a distribuição a posteriori para o total populacional Λ .

Resultado 4.2. Suponha que o vetor de valores populacionais $X = (X_1, ..., X_N)$ assumindo valores em \mathbb{N}^N é tal que vale a seguinte condição de multinomialidade:

$$P(X_1 = x_1, \dots, X_N = x_N | \Lambda = t) = {t \choose x_1 \dots x_N} N^{-t} I_{\{t\}}(\sum_{i=1}^N x_i) .$$

Considerando $P(\Lambda = t) = p_t$, t = 0, 1, ..., como a incerteza a priori sobre o total populacional Λ , então sua distribuição a posteriori, sob a abordagem preditivista para populações finitas, é dada por:

$$P\left(\Lambda = t \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \binom{t}{\sum_{i=1}^n x_i} \left(\frac{n}{N}\right)^{\sum_{i=1}^n x_i} \left(1 - \frac{n}{N}\right)^{t - \sum_{i=1}^n x_i} p_t, \quad (4.4)$$

para $t \ge \sum_{i=1}^{n} x_i$. Para $t < \sum_{i=1}^{n} x_i$, $P(\Lambda = t | X_1 = x_1, \dots, X_n = x_n) = 0$.

Demonstração:

Seja $\boldsymbol{x} = (x_1, \dots, x_n) \operatorname{com} x_i \in \mathbb{N}, i = 1, \dots, n$, e considere o conjunto $S = \{(y_{n+1}, \dots, y_N) \in \mathbb{N}\}$

$$\begin{split} \mathbb{N}^{N-n} &: \sum_{j=n+1}^{N} y_j = t - \sum_{i=1}^{n} x_i \}. \text{ Assim,} \\ P\Big(X_1 = x_1, \dots, X_n = x_n \Big| \Lambda = t \Big) = \\ &= \sum_{S} P\Big(X_1 = x_1, \dots, X_n = x_n, X_{n+1} = y_{n+1}, \dots, X_N = y_N \Big| \Lambda = t \Big) \\ &= \Big(\frac{1}{N}\Big)^t \sum_{S} \frac{t!}{x_1! \dots x_n! y_{n+1}! \dots y_N!} \\ &= \Big(\frac{1}{N}\Big)^t \frac{t!}{x_1! \dots x_n!} \sum_{S} \frac{1}{y_{n+1}! \dots y_N!} \\ &= \Big(\frac{1}{N}\Big)^t \frac{t!}{x_1! \dots x_n! (t - \sum_{i=1}^{n} x_i)!} \sum_{S} \frac{(t - \sum_{i=1}^{n} x_i)!}{y_{n+1}! \dots y_N!} \\ &= \frac{1}{x_1! \dots x_n!} \frac{t!}{(t - \sum_{i=1}^{n} x_i)!} \frac{(N-n)^{t - \sum_{i=1}^{n} x_i}}{N^t} \\ &= \frac{(\sum_{i=1}^{n} x_i)!}{x_1! \dots x_n!} \Big(\frac{1}{n}\Big)^{\sum_{i=1}^{n} x_i} \Big(\frac{n}{N}\Big)^{\sum_{i=1}^{n} x_i} \Big(\frac{N-n}{N}\Big)^{t - \sum_{i=1}^{n} x_i} . \end{split}$$

A distribuição a posteriori de Λ é então dada por:

$$P\left(\Lambda = t \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto$$

$$\propto P\left(X_1 = x_1, \dots, X_n = x_n \middle| \Lambda = t\right) p_t$$

$$= \frac{\left(\sum_{i=1}^n x_i\right)!}{x_1! \dots x_n!} \left(\frac{1}{n}\right)^{\sum_{i=1}^n x_i} \left(\frac{t}{\sum_{i=1}^n x_i}\right) \left(\frac{n}{N}\right)^{\sum_{i=1}^n x_i} \left(1 - \frac{n}{N}\right)^{t - \sum_{i=1}^n x_i} p_t$$

$$\propto \left(\frac{t}{\sum_{i=1}^n x_i}\right) \left(\frac{n}{N}\right)^{\sum_{i=1}^n x_i} \left(1 - \frac{n}{N}\right)^{t - \sum_{i=1}^n x_i} p_t.$$

Note que a distribuição a posteriori para o total populacional nesta situação é proporcional ao produto da verossimilhança de uma variável aleatória com distribuição Binomial de parâmetros $t \in \frac{n}{N}$ pela priori e se concentra em $\{\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i + 1, ...\}$. Note ainda que a posteriori depende da amostra apenas através da estatística suficiente $\sum_{i=1}^{n} X_i$ e pode ser facilmente mostrado que a distribuição em (4.4) coincide com a distribuição a posteriori de Λ dado $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i$. No entanto, a priori, Λ pode assumir qualquer valor no conjunto dos naturais. Resta então especificar a distribuição a priori para Λ . Para isto, consideramos o seguinte resultado envolvendo as distribuições Exponencial e Geométrica:

Resultado 4.3 (Relação entre distribuições Exponencial e Geométrica). Seja X uma variável aleatória tal que X ~ Exponencial(b), b > 0. Então, a variável $Y = \lfloor NX \rfloor$ segue distribuição Geométrica de parâmetro $1 - e^{-\frac{b}{N}}$, isto é,

$$P(Y = y) = e^{-\frac{b}{N}y}(1 - e^{-\frac{b}{N}}), \quad y = 0, 1, 2, \dots$$

A distribuição Geométrica possui suporte no conjunto dos naturais, o que a torna uma boa "candidata"como distribuição a priori para Λ . Além disso, como apresentado no resultado 4.3, ela é equivalente a uma discretização da distribuição Exponencial (que será utilizada no modelo de superpopulação).

A distribuição em (4.4) é a distribuição a posteriori para Λ sob a abordagem preditivista para populações finitas. Estendendo este modelo para uma sequência infinita de variáveis aleatórias, estas variáveis, dado λ , são independentes e identicamente distribuídas segundo a distribuição de Poisson, isto é, $(X_n)_{n\geq 1}|\lambda \stackrel{ciid}{\sim} Poisson(\lambda)$ (Wechsler, 1993).

Desta forma, podemos formular o modelo de superpopulação da seguinte maneira: representamos nossa incerteza a priori sobre λ através da distribuição Gama(a, b), a, b > 0. Mais especificamente, podemos utilizar a Gama(1, b) (*Exponencial(b)*), de modo que $E(\lambda) = 1/b$. A distribuição Gama é bastante flexível, podendo assumir diversas formas conforme variamos os seus parâmetros e possui suporte em \mathbb{R}_+ , o que a torna uma boa escolha como distribuição a priori para λ . Além disso, a distribuição Gama é conjugada à Poisson, o que leva à posteriori: $\lambda |\sum_{i=1}^{n} x_i \sim Gama(1 + \sum_{i=1}^{n} x_i, b + n)$.

Novamente, o interesse está em testar hipóteses unilaterais e simples através das abordagens preditivista e de superpopulação. Para isto, consideramos as hipóteses:

$$\begin{cases}
H_0: \lambda \le \lambda_0 \\
H_1: \lambda > \lambda_0
\end{cases}$$
(4.5)

е

$$\begin{cases} H'_0 : \lambda = \lambda_0 \\ H'_1 : \lambda \neq \lambda_0 \end{cases}, \tag{4.6}$$

,

 $\lambda_0 > 0$. Consideramos então, a priori, a distribuição Gama para λ , no modelo de superpopulação, e a distribuição Geométrica, que é "análoga" à Gama no caso discreto como mencionado no Resultado 4.3, para Λ no modelo de populações finitas. Desta forma, comparamos os resultados para diferentes médias das distribuições a priori.

Considerando $\lambda_0 = 3, 5$, a probabilidade da hipótese nula $H_0 : \lambda \leq 3, 5$, sob a abordagem de superpopulação, é dada por

$$P\left(H_0 \left| \sum_{i=1}^n X_i = s \right) = \int_0^{3,5} \frac{(b+n)^{1+s}}{\Gamma(1+s)} \lambda^s e^{-(b+n)s} d\lambda.$$

Por outro lado, no modelo de populações finitas, esta probabilidade é

$$P\left(H_0\left|\sum_{i=1}^n X_i=s\right) = \frac{1}{c}\sum_{t=s}^{\lfloor 3,5N \rfloor} \binom{t}{s} \left(\frac{n}{N}\right)^s \left(1-\frac{n}{N}\right)^{t-s} e^{-\frac{b}{N}t} (1-e^{-\frac{b}{N}}),$$

sendo $c = \sum_{t=s}^{\infty} P(X_1 = x_1, \dots, X_n = x_n | \Lambda = t) p_t$ a constante normalizadora.

No caso da hipótese precisa $H_0': \lambda = 3, 5$, a região tangente e a evidência em favor de H_0' , sob o enfoque de superpopulação, são

$$T_{x} = \left\{ \lambda > 0 : f(\lambda|s) > f(\lambda_{0}|s) \right\}$$

e
$$ev(\{\lambda_{0}\}; \boldsymbol{x}) = 1 - \int_{T_{x}} \frac{(b+n)^{1+s}}{\Gamma(1+s)} \lambda^{s} e^{-(b+n)s} d\lambda$$

sendo f(.|s) a função densidade da distribuição a posteriori Gama(1+s, b+n). Por outro

lado, considerando $\Lambda_0 = N\lambda_0 \in \mathbb{N}$, sob o enfoque preditivista para populações finitas, temos

$$T_{x} = \left\{ k \in \mathbb{N} : P\left(\Lambda = k \left| \sum_{i=1}^{n} X_{i} = s \right) > P\left(\Lambda = \Lambda_{0} \left| \sum_{i=1}^{n} X_{i} = s \right) \right\} \right\}$$

e
$$ev\left(\{\Lambda_{0}\}; \boldsymbol{x}\right) = 1 - \sum_{T_{x}} P\left(\Lambda = k \left| \sum_{i=1}^{n} X_{i} = s \right) \right]$$

Nas Figuras 4.8-4.10, observamos que, quando o tamanho amostral é grande, a diferença entre os métodos se sobressai, sendo que a diferença máxima entre as probabilidades da hipótese nula dos testes unilaterais chega a 22%, quando consideramos uma distribuição a priori para λ (Λ) com média 1 (100) e a média amostral é 3,65. Quando a média da priori para λ é 3,5, essa diferença diminui, chegando a no máximo 18%, quando a média amostral é 3,4, por exemplo. Comparando os resultados obtidos utilizando prioris com médias 1 e 10 com os resultados obtidos quando a média da priori é 3,5 (caso em que a média da priori é igual a λ_0), observamos que os efeitos das prioris sobre as probabilidades da hipótese nula a posteriori são similares nos dois métodos quando a amostra é pequena, e diminuem para amostras maiores. Contudo, quando a amostra é grande, os resultados obtidos pelo modelo de superpopulação são mais afetados pelas distribuições a priori do que os resultados do modelo preditivista para populações finitas .

As Figuras 4.11-4.13 apresentam os valores da evidência em favor de H'_0 : $\lambda = 3, 5$. Assim como no caso de hipóteses unilaterais, observamos um comportamento similar entre as evidências calculadas segundo as duas abordagens quando o tamanho amostral é pequeno e as diferenças se destacam mais conforme o tamanho amostral cresce em relação ao populacional. Por exemplo, considerando a priori para λ com média 3,5 (Figura 4.12(b)), quando a média amostral é 3,275, a evidência sob o enfoque de superpopulação é 0,25 enquanto que sob a abordagem preditivista este valor cai para 0,02, considerando amostra de tamanho 80. Para n = 20, a diferença máxima é de 0,06, quando a média amostral é 4. O efeito das distribuições a priori sobre os valores das evidências é maior no modelo de superpopulação do que no modelo de populações finitas e as prioris influenciam mais os resultados obtidos para amostras pequenas.



Análises considerando outros valores de $N \in \lambda_0$ são apresentadas no Apêndice B.

Figura 4.8: Valores da probabilidade a posteriori de $H_0: \lambda \leq 3, 5 \ (\Lambda \leq 3, 5N)$ em função da média amostral para uma população de tamanho $N = 100 \ e \ E(\lambda) = 1 \ (E(\Lambda) \approx 100)$.



Figura 4.9: Valores da probabilidade a posteriori de $H_0: \lambda \leq 3, 5 \ (\Lambda \leq 3, 5N)$ em função da média amostral para uma população de tamanho $N = 100 \ e \ E(\lambda) = 3, 5 \ (E(\Lambda) \approx 350)$.



Figura 4.10: Valores da probabilidade a posteriori de H_0 : $\lambda \leq 3, 5$ ($\Lambda \leq 3, 5N$) em função da média amostral para uma população de tamanho $N = 100 \ e \ E(\lambda) = 10 \ (E(\Lambda) \approx 1000)$.



Figura 4.11: Valores da evidência em favor de H'_0 : $\lambda = 3,5$ ($\Lambda = 3,5N$) em função da média amostral para uma população de tamanho N = 100 e $E(\lambda) = 1$ ($E(\Lambda) \approx 100$).



Figura 4.12: Valores da evidência em favor de H'_0 : $\lambda = 3, 5$ ($\Lambda = 3, 5N$) em função da média amostral para uma população de tamanho N = 100 e $E(\lambda) = 3, 5$ ($E(\Lambda) \approx 350$).



Figura 4.13: Valores da evidência em favor de H'_0 : $\lambda = 3, 5$ ($\Lambda = 3, 5N$) em função da média amostral para uma população de tamanho N = 100 e $E(\lambda) = 10$ ($E(\Lambda) \approx 1000$).

4.3 Distribuição Uniforme em \mathbb{N}

Nesta seção, consideramos novamente que o vetor de valores populacionais desconhecidos, $X = (X_1, \ldots, X_N)$, assume valores no conjunto \mathbb{N}^N . No entanto, nosso interesse está no maior valor populacional, $X_{(N)} = max\{X_1, \ldots, X_N\}$, e não mais no total (média) populacional. Além disso, a suposição de multinomialidade (apresentada na seção anterior) é substituída por uma condição de uniformidade, como apresentado a seguir:

Resultado 4.4. Suponha que a distribuição de $X_1, \ldots, X_N | X_{(N)} = t, t \in \mathbb{N}$, é uniforme no conjunto $A_t = \{(x_1, \ldots, x_N) \in \mathbb{N}^N : max\{x_1, \ldots, x_N\} = t\}$ e considere $P(X_{(N)} = j) = p_j$, $j = 0, 1, \ldots$, a incerteza a priori para o máximo populacional. A função de probabilidade a posteriori de $X_{(N)}$, dada uma amostra x_1, \ldots, x_n , é então dada por:

$$P\left(X_{(N)} = t \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \begin{cases} \frac{1}{(t+1)^n} \left(\frac{1}{1-\left(\frac{t}{t+1}\right)^N}\right) p_t, se \ max\{x_1, \dots, x_n\} = t \\ \frac{1}{(t+1)^n} \left(\frac{1-\left(\frac{t}{t+1}\right)^{N-n}}{1-\left(\frac{t}{t+1}\right)^N}\right) p_t, se \ max\{x_1, \dots, x_n\} < t \end{cases}$$

$$(4.7)$$

Além disso, $P(X_{(N)} = t | X_1 = x_1, \dots, X_n = x_n) = 0$ se $max\{x_1, \dots, x_n\} > t$.

Demonstração: Seja o conjunto $A_m = \{(x_1, \dots, x_N) \in \mathbb{N}^N : max\{x_1, \dots, x_N\} = m\},$ $m \in \mathbb{N}$. Assim, a partir da hipótese de uniformidade, temos

$$P(X_1 = x_1, \dots, X_N = x_N | X_{(N)} = m) = \frac{1}{\#(A_m)} I_{\{m\}}(max\{x_1, \dots, x_N\}) ,$$

sendo #(A) a quantidade de elementos do conjunto A. Para calcularmos $\#(A_m)$, considere os conjuntos $M_i = \{(x_1, \ldots, x_N) \in A_m : x_i = m\}, i = 1, \ldots, N$, de tal forma que $A_m = \bigcup_{i=1}^N M_i$. Então,

$$#(A_m) = \#\left(\bigcup_{i=1}^N M_i\right) = \sum_{i=1}^N \#(M_i) - \sum_{1 \le i < j \le N} \#(M_i \cap M_j) + \sum_{1 \le i < j < k \le N} \#(M_i \cap M_j \cap M_k) + \dots + (-1)^{N+1} \#(M_1 \cap M_2 \cap \dots \cap M_N) .$$

Como $\#(M_{l_1} \cap M_{l_2} \cap \ldots \cap M_{l_k}) = (m+1)^{N-k}$ para $\{l_1, \ldots, l_k\} \subset \{1, \ldots, N\}$, temos

$$\#(A_m) = \sum_{i=1}^N \binom{N}{i} (-1)^{i+1} (m+1)^{N-i} = (m+1)^N - m^N$$

e, portanto,

$$P(X_1 = x_1, \dots, X_N = x_N | X_{(N)} = m) = \frac{1}{(m+1)^N - m^N} I_{\{m\}}(max\{x_1, \dots, x_N\}) .$$

Consideramos agora a amostra $\boldsymbol{x} = (x_1, \ldots, x_n), x_i \in \mathbb{N}, i = 1, \ldots, n, e$ o conjunto $B = \{(y_{n+1}, \ldots, y_N) \in \mathbb{N}^{N-n} : max\{x_1, \ldots, x_n, y_{n+1}, \ldots, y_N\} = m\}.$ A probabilidade da amostra \boldsymbol{x} , dado o máximo populacional $X_{(N)}$ é

$$P(X_1 = x_1, \dots, X_n = x_n | X_{(N)} = m) =$$

= $\sum_B P(X_1 = x_1, \dots, X_n = x_n, X_{n+1} = y_{n+1}, \dots, X_N = y_N | X_{(N)} = m).$

Desta forma, se $x_{(n)} = max\{x_1, \ldots, x_n\} < m$, então $B = \{(y_{n+1}, \ldots, y_N) \in \mathbb{N}^{N-n} : max\{y_{n+1}, \ldots, y_N\} = m\}$, que é similar ao conjunto A_m , só que considerando vetores com dimensão N - n. A distribuição a posteriori é, nesse caso,

$$P\left(X_{(N)} = m \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \frac{(m+1)^{N-n} - m^{N-n}}{(m+1)^N - m^N} p_m I_{\{0,1,\dots,m-1\}}(max\{x_1,\dots,x_n\})$$

Por outro lado, se $x_{(n)} = max\{x_1, \ldots, x_n\} = m$, então $B = \{(y_{n+1}, \ldots, y_N) \in \mathbb{N}^{N-n} : max\{y_{n+1}, \ldots, y_N\} \le m\}$ e a posteriori é dada por:

$$P\left(X_{(N)} = m \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \frac{(m+1)^{N-n}}{(m+1)^N - m^N} p_m I_{\{m\}}(max\{x_1, \dots, x_n\})$$

Agrupando os dois casos, temos então que a distribuição a posteriori de $X_{(N)}$ é dada por:

$$P\left(X_{(N)} = t \middle| X_1 = x_1, \dots, X_n = x_n\right) \propto \begin{cases} \frac{1}{(t+1)^n} \left(\frac{1}{1-\left(\frac{t}{t+1}\right)^N}\right) p_t, se \ x_{(n)} = t \\ \frac{1}{(t+1)^n} \left(\frac{1-\left(\frac{t}{t+1}\right)^{N-n}}{1-\left(\frac{t}{t+1}\right)^N}\right) p_t, se \ x_{(n)} < t \end{cases}$$

A distribuição em (4.7) é a distribuição a posteriori para o máximo de um conjunto finito de variáveis que assumem valores naturais sob a abordagem preditivista para populações finitas.

Outra forma de inferirmos sobre o máximo populacional, considerando um modelo de superpopulação, consiste em supor que a sequência de variáveis aleatórias $(X_n)_{n\geq 1}$, dado θ , seguem distribuição Uniforme Discreta, isto é $(X_n)_{n\geq 1}|\theta \stackrel{ciid}{\sim} U\{0, 1, \ldots, \theta\}, \theta \in \mathbb{N}$. Desta forma, o interesse está no parâmetro θ , que representa o valor de $X_{(N)}$ quando $N \to \infty$. Nesta situação, definindo $P(\theta = k) = p_k$, k=0,1,..., a incerteza a priori, temos a seguinte distribuição a posteriori para o parâmetro θ

$$P\left(\theta = k \middle| X_1 = x_1, \dots, X_n = x_n\right) = \frac{1}{(k+1)^n} p_k I_{\{x_{(n)}, x_{(n)}+1, \dots\}}(k) , \qquad (4.8)$$

sendo $x_{(n)} = max\{x_1, \ldots, x_n\}$. Note que, ao contrário do que ocorre nos modelos Bernoulli e Poisson, no modelo Uniforme a distribuição a priori para o máximo populacional sob o enfoque preditivista para populações finitas pode ser a mesma utilizada no modelo de superpopulação, pois o espaço paramétrico é o mesmo, N. Contudo, não é necessário que estas distribuições sejam iguais.

Nesse caso, o principal interesse está em encontrar o valor máximo da população. Desse modo, testaremos apenas hipóteses unilaterais, pois o resultado deste teste já permite-nos inferir sobre θ . Testamos, então, as hipóteses:

$$\begin{cases} H_0: \theta \le \theta_0 \\ H_1: \theta > \theta_0 \end{cases}, \tag{4.9}$$

para $\theta_0 \in \mathbb{N}$, através dos modelos preditivista para populações finitas e de superpopulação.

Para realizarmos os testes de hipóteses e compararmos os resultados produzidos pelos dois métodos, consideramos, a priori, uniformidade em um subconjunto finito de N da forma $\{0, 1, \ldots, M\}, M >> \theta$, não privilegiando nenhum ponto desse conjunto. Atribuímos, então, diferentes valores para o tamanho da população N, o tamanho amostral n, o máximo populacional $X_{(N)}$ e o valor que queremos testar θ_0 . Para cada combinação dessas quantidades, consideramos todos os possíveis resultados da estatística suficiente $X_{(n)} = max\{X_1, \ldots, X_n\},$ n < N, e, para cada um destes resultados, é calculada a probabilidade a posteriori da hipótese nula H_0 .

Nas figuras 4.14-4.21, são apresentadas as probabilidade da hipótese nula em função do máximo amostral para alguns valores fixados de N, n, $X_{(N)} \in \theta_0$. Observamos que as maiores diferenças entre os resultados das duas abordagens ocorrem nas situações em que o tamanho amostral é grande (em relação ao populacional) e o máximo populacional não é muito menor que o tamanho populacional (Figuras 4.15(a), 4.15(b), 4.19(a), 4.19(b), 4.21(a) e 4.21(b)). Por exemplo, no caso da Figura 4.19(a), as probabilidades de $H_0: \theta \leq 90$ são 0,58 e 0,88, sob os métodos de superpopulação e população finita, respectivamente, quando $X_{(n)} = 90$. Para a Figura 4.15(b), essas probabilidades são 0,37 e 0,73, quando $X_{(n)} = 40$.

Nas amostras pequenas, as diferença são menores, mas podem chegar a mais de 10%, conduzindo à conclusões distintas. Por exemplo, na Figura 4.20(a), as probabilidades de H_0 sob os modelos de superpopulação e preditivista são 0,09 e 0,23, respectivamente, quando $X_{(n)} = 180$. Nas situações em que o máximo populacional é muito menor que a quantidade de unidades populacionais (Figuras 4.16 e 4.17), quase não há diferença entre os métodos, independentemente do tamanho da amostra. As diferenças são muito pequenas também quando o valor de θ_0 é maior que o máximo populacional $X_{(N)}$, situação na qual as probabilidades ficam sempre muito próximas de 1 nos dois modelos avaliados.



Figura 4.14: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 30, $n = 10, X_{(N)} = 40$.



Figura 4.15: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 30, n = 20, $X_{(N)} = 40$.



Figura 4.16: Valores da probabilidade a posteriori de $H_0: \theta \leq \theta_0$ $(X_{(N)} \leq \theta_0)$ para N = 100, $n = 20, X_{(N)} = 40$.



Figura 4.17: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 100, n = 80, $X_{(N)} = 40$.



Figura 4.18: Valores da probabilidade a posteriori de $H_0: \theta \leq \theta_0$ $(X_{(N)} \leq \theta_0)$ para N = 100, $n = 20, X_{(N)} = 100$.



Figura 4.19: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 100, $n = 80, X_{(N)} = 100$.



Figura 4.20: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 100, $n = 20, X_{(N)} = 200$.



Figura 4.21: Valores da probabilidade a posteriori de H_0 : $\theta \leq \theta_0$ ($X_{(N)} \leq \theta_0$) para N = 100, $n = 80, X_{(N)} = 200$.

4.4 Distribuição Multinomial

O modelo Bernoulli, apresentado na primeira seção deste capítulo, descreve uma situação na qual a característica de cada unidade populacional pode assumir somente dois possíveis valores, representados, numericamente, por 0 ou 1. Contudo, este modelo pode ser generalizado para uma versão na qual cada variável assume um entre M possíveis resultados, representados pelos valores a_1, \ldots, a_M . Neste caso, consideramos que o vetor $\mathbf{X} = (X_1, \ldots, X_N)$ é permutável e que X_i assume valores no conjunto $A = \{a_1, \ldots, a_M\}, a_i \in \mathbb{R}, i = 1, \ldots, N$.

Nesta situação, o interesse é fazer inferência sobre o vetor de totais populacionais $T(X) = (T_1(X), \ldots, T_M(X))$, sendo $T_j(X)$ o total de elementos na população para os quais a característica de interesse assume o valor a_j , isto é, $T_j(X) = \sum_{i=1}^N I_{\{a_j\}}(X_i), j = 1, \ldots, N$. Para isso, definimos os vetores $Y_i = (I_{\{a_1\}}(X_i), \ldots, I_{\{a_M\}}(X_i)), i = 1, \ldots, N$, isto é, cada vetor Y_i tem dimensão M e apresenta o valor 1 na posição k e 0 nas demais posições, quando $X_i = a_k$, $k = 1, \ldots, M, i = 1, \ldots, N$. Desta forma, Y_1, \ldots, Y_N são vetores aleatórios permutáveis (Iglesias, 1993). Além disso, temos:

$$P(X_1 = x_1, \dots, X_N = x_N) = P(Y_1 = y_1, \dots, Y_N = y_N),$$
 (4.10)

sendo $\boldsymbol{y}_{\boldsymbol{i}} = (I_{\{a_1\}}(x_i), \dots, I_{\{a_M\}}(x_i)), \ i = 1, \dots, N.$

Uma maneira simples de pensarmos neste modelo é imaginarmos uma urna com N bolas em que cada bola está rotulada com algum dos números a_1, \ldots, a_M diferentes. O interesse está em descobrirmos os totais de bolas de cada tipo dentro da urna, ou seja, o vetor $T = \sum_{i=1}^{N} Y_i = (T_1, \ldots, T_M)$. O resultado a seguir apresenta a distribuição a posteriori para T segundo a abordagem preditivista para populações finitas, considerando uma amostra $(y_1, \ldots, y_n), n < N$.

Resultado 4.5. Sejam os vetores Y_1, \ldots, Y_N permutáveis descritos anteriormente e considere $P(T = t) = p_t, t \in \{(t_1, \ldots, t_M) \in \mathbb{N}^M : \sum_{i=1}^M t_i = N\}$, como a incerteza a priori a respeito do vetor **T**. Então, sua função de probabilidade a posteriori é dada por:

$$P\left(\boldsymbol{T}=\boldsymbol{t}\middle|\boldsymbol{Y}_{1}=\boldsymbol{y}_{1},\ldots,\boldsymbol{Y}_{n}=\boldsymbol{y}_{n}\right)\propto\frac{\binom{N-n}{t_{1}-t_{1s}\ldots t_{M}-t_{Ms}}}{\binom{N}{t_{1}\ldots t_{M}}}p_{t}\prod_{i=1}^{M}I_{\{t_{is},\ldots,N-\sum_{j\neq i}t_{js}\}}(t_{i}),$$

sendo $\mathbf{t} = (t_1, \dots, t_M) \in \mathbb{N}^M$ com $\sum_{i=1}^M t_i = N$ e $\mathbf{t}_s = \sum_{i=1}^n \mathbf{y}_i = (t_{1s}, \dots, t_{Ms}) \in \mathbb{N}^M$ com $\sum_{i=1}^M t_{is} = n.$

Demontração: Da hipótese de permutabilidade, temos (Iglesias, 1993):

$$P\left(\boldsymbol{Y_1} = \boldsymbol{y_1}, \dots, \boldsymbol{Y_N} = \boldsymbol{y_N} \middle| \boldsymbol{T} = \boldsymbol{t}\right) = \frac{1}{\binom{N}{t_1 \dots t_M}} I_{\{\boldsymbol{t}\}}(\sum_{i=1}^N \boldsymbol{y_i}) \ .$$

Sejam $y = (y_1, \dots, y_n), y_i \in \{e_1, \dots, e_M\}$, onde $e_j = (0, \dots, 0, \underbrace{1}_{j}, 0, \dots, 0), j = 1, \dots, M$, e o conjunto $S = \{(v_{n+1}, \dots, v_N) \in \{e_1, \dots, e_M\}^{N-n} : \sum_{j=n+1}^{N} v_j = t - t_s\}$. Assim,

$$P\left(\mathbf{Y_1} = \mathbf{y_1}, \dots, \mathbf{Y_n} = \mathbf{y_n} \middle| \mathbf{T} = \mathbf{t}\right) =$$

$$= \sum_{S} P\left(\mathbf{Y_1} = \mathbf{y_1}, \dots, \mathbf{Y_n} = \mathbf{y_n}, \mathbf{Y_{n+1}} = \mathbf{v_{n+1}}, \dots, \mathbf{Y_N} = \mathbf{v_N} \middle| \mathbf{T} = \mathbf{t}\right)$$

$$= \sum_{S} \frac{1}{\binom{N}{t_1 \dots t_M}} = \frac{\binom{N-n}{t_1 - t_{1s} \dots t_M - t_{Ms}}}{\binom{N}{t_1 \dots t_M}}.$$

A distribuição a posteriori de T é então dada por:

$$P(\mathbf{T} = \mathbf{t} | \mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_n = \mathbf{y}_n) \propto$$

$$\propto P(\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_n = \mathbf{y}_n | \mathbf{T} = \mathbf{t}) p_t$$

$$= \frac{\binom{N-n}{t_1 - t_{1s} \dots t_M - t_{Ms}}}{\binom{N}{t_1 \dots t_M}} p_t,$$

para $t_{is} \leq t_i \leq N - \sum_{j \neq i} t_{js}, i = 1, \dots, M.$

A distribuição apresentada no Resultado 4.5 representa a distribuição a posteriori do vetor de totais populacionais T, dado a amostra (y_1, \ldots, y_n) . Não é difícil verificar que esta distribuição coincide com a distribuição de T dado o vetor de totais da amostra, $\sum_{i=1}^{n} Y_i = t_s$.

Do ponto de vista bayesiano de superpopulação, modelamos não os totais, mas as proporções populacionais, isto é, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{M-1})$, onde π_i representa a proporção de elementos populacionais para os quais a característica de interesse vale a_i , $i = 1, \ldots, M - 1$, de modo que $\pi_i > 0$ e $\pi_1 + \ldots + \pi_{M-1} < 1$. Assim, supomos que a incerteza a priori sobre $\boldsymbol{\pi}$ pode ser modelada através de uma distribuição Dirichlet de ordem M - 1, $\boldsymbol{\pi} \sim Dir(\boldsymbol{\alpha})$, com $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$, $\alpha_i > 0$, $i = 1, \ldots, M$. Supomos ainda que a sequência de vetores aleatórias $(\boldsymbol{Y_n})_{n\geq 1}$, dado $\boldsymbol{\pi}$, seguem distribuição Multinomial: $(\boldsymbol{Y_n})_{n\geq 1}|\boldsymbol{\pi} \stackrel{ciid}{\sim} Mult(1, \boldsymbol{\pi})$. Desta forma, a distribuição a posteriori de $\boldsymbol{\pi}$ também será Dirichlet, com os parâmetros atualizados, isto é, $\boldsymbol{\pi}|\sum_{i=1}^n \boldsymbol{Y_i} = \boldsymbol{t_s} \sim Dir(\boldsymbol{\alpha} + \boldsymbol{t_s})$, e sua função densidade é dada por:

$$f(\boldsymbol{\pi}|\boldsymbol{t_s}) = \frac{1}{\beta(\boldsymbol{\alpha} + \boldsymbol{t_s})} \prod_{i=1}^{M} \pi_i^{\alpha_i + t_{is} - 1}, \qquad (4.11)$$

para $\pi_1, \ldots, \pi_{M-1} > 0$, $\sum_{i=1}^{M-1} \pi_i < 1$, $\pi_M = 1 - \sum_{i=1}^{M-1} \pi_i \in \beta(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^M \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^M \alpha_i)}$.

Considerando o modelo Multinomial descrito acima com M = 3, nossa análise contemplará dois problemas: o primeiro consiste em comparar duas proporções $\pi_1 \in \pi_2$, de tal modo que testaremos se uma proporção é menor ou igual a outra, $H_0 : \pi_1 \leq \pi_2$, ou simplesmente se há igualdade entre essas proporções, $H'_0 : \pi_1 = \pi_2$. O segundo problema consiste em testar simultaneamente se as 3 proporções são menores que um certo valor fixo $\pi_0 \in (0, 1)$, de tal forma que a hipótese nula é $H_0 : \bigcap_{i=1}^3 \{\pi_i \leq \pi_0\}$.

Para estas análises, consideramos que, a priori, não temos informações para privilegiar nenhum ponto do espaço paramétrico e, portanto, utilizamos distribuições uniformes multivariadas. Desse modo, modelamos a incerteza a priori através das seguintes distribuições:

T ~ Unif(A), onde A = {(t₁, t₂, t₃) ∈ N³ : ∑_{i=1}³ t_i = N}, sob a abordagem preditivista para populações finitas;

• $\boldsymbol{\pi} \sim Dir(1, 1, 1)$, sob o modelo de superpopulação.

4.4.1 Comparação entre duas proporções

Considerando o modelo multinomial com 3 categorias (M=3), podemos avaliar a relação entre duas das 3 proporções envolvidas. Inicialmente, consideramos o caso de hipóteses compostas não precisas:

$$\begin{cases}
H_0: \pi_1 \le \pi_2 \\
H_1: \pi_1 > \pi_2
\end{cases}$$
(4.12)

Para estas hipóteses, utilizaremos o teste de Bayes baseado em probabilidades a posteriori.

Neste caso, sob o enfoque bayesiano de superpopulações, a probabilidade a posteriori da hipótese nula $H_0:\pi_1\leq\pi_2$ é

$$P\left(H_0 \middle| \sum_{i=1}^n \mathbf{Y}_i = \mathbf{t}_s \right) = \int_0^{0.5} \int_{\pi_1}^{1-\pi_1} \frac{1}{\beta(\alpha + \mathbf{t}_s)} \prod_{i=1}^3 \pi_i^{\alpha_i + t_{is} - 1} d\pi_2 d\pi_1 , \qquad (4.13)$$

onde $\pi_3 = 1 - \pi_1 - \pi_2$, enquanto que, sob a abordagem preditivista para populações finitas, a hipótese nula equivalente é $H_0: T_1 \leq T_2$ e sua probabilidade é dada por:

$$P\left(H_0 \middle| \sum_{i=1}^n \mathbf{Y}_i = \mathbf{t}_s \right) = \frac{1}{c} \sum_{t_2=t_{2s}}^{N-t_{1s}-t_{3s}} \sum_{t_1=t_{1s}}^{t_2} \frac{\binom{N-n}{t_1-t_{1s} t_2 - t_{2s} t_3 - t_{3s}}}{\binom{N}{t_1 t_2 t_3}} p_t \prod_{i=1}^3 I_{\{t_{is},\dots,N-\sum_{j\neq i} t_{js}\}}(t_i)$$

sendo $c = \sum_{\boldsymbol{t} \in D} P(\boldsymbol{Y}_1 = \boldsymbol{y}_1, \dots, \boldsymbol{Y}_n = \boldsymbol{y}_n | \boldsymbol{T} = \boldsymbol{t}) p_t$, onde $D = \{(t_1, t_2, t_3) \in \mathbb{N}^3 : t_1 + t_2 + t_3 = N \text{ e } t_{is} \leq t_i \leq N - \sum_{j \neq i} t_{js}, i = 1, 2, 3\}.$

No modelo de superpopulação, a integral em (4.13) não tem forma analítica simples e, desse modo, optamos por calculá-la através de métodos numéricos, utilizando simulação Monte Carlo. Para isto, o primeiro passo é a simulação de uma amostra com 10⁴ elementos da posteriori Dirichlet. A probabilidade a posteriori da hipótese nula é então estimada pela proporção de elementos amostrados nos quais π_1 é menor ou igual a π_2 . Por outro lado, no caso de populações finitas, a probabilidade de H_0 é obtida somando todas as probabilidades a posteriori dos pontos em que T_1 é menor ou igual a T_2 . A Figura 4.22 exibe as probabilidades a posteriori de H_0 considerando as abordagens preditivista para populações finitas e de superpopulação para uma população de tamanho N = 100 e tamanhos amostrais n = 20 e n = 80, como função da estatística suficiente (t_{1s}, t_{2s}) (nesse caso, equivalente a $\mathbf{t}_s = \sum_{i=1}^n \mathbf{y}_i = (t_{1s}, t_{2s}, t_{3s})$). Observamos que, para as amostras com tamanho 20, as superfícies (representações gráficas das probabilidades) formadas para as duas abordagens são bem similares. Neste caso, as maiores diferenças entre as probabilidades a posteriori de H_0 são de pouco mais de 5% e ocorrem para $\mathbf{t}_s = (0, 0, 20)$ e $\mathbf{t}_s = (0, 1, 19)$.

Por outro lado, quando aumentamos o tamanho amostral para 80, o comportamento das curvas é diferente, sendo mais suave no caso dos valores calculados sob a abordagem de superpopulação. No caso de populações finitas, os valores das probabilidades aumentam de forma mais "rápida". As diferenças nas probabilidades a posteriori chegam a 33% quando $\mathbf{t}_{s} = (0, 0, 80)$, mas ficam em torno de 20% para uma grande quantidade de possíveis amostras. Por exemplo, se $\mathbf{t}_{s} = (37, 31, 12)$, então as probabilidades são 0,06 e 0,23, sob as abordagens de populações finitas e superpopulação, respectivamente. Para $\mathbf{t}_{s} = (39, 33, 8)$, essas probabilidades são 0,07 e 0,24.

Agora, considere que estamos interessados em saber se duas proporções são diferentes. Neste sentido, testamos as seguintes hipóteses:

$$\begin{cases} H'_0: \pi_1 = \pi_2 \\ H'_1: \pi_1 \neq \pi_2 \end{cases}$$
 (4.14)

Como a hipótese nula descrita em (4.14) é uma hipótese precisa, utilizaremos o FBST para a realização do teste. Consideramos, primeiramente, o caso de superpopulação. O primeiro passo é encontrar o valor máximo da função densidade Dirichlet a posteriori sob a hipótese nula que, neste caso, pode ser reescrita como

$$f_0(\pi_1 | \boldsymbol{t_s}) = \frac{1}{\beta(\boldsymbol{\alpha} + \boldsymbol{t_s})} \pi_1^{\alpha_1 + \alpha_2 + t_{1s} + t_{2s} - 2} (1 - 2\pi_1)^{\alpha_3 + t_{3s} - 1} I_{(0, 0, 5)}(\pi_1) .$$
(4.15)

A função $f_0(.|\boldsymbol{t}_s)$ atinge o valor máximo em $\pi_1^* = \frac{\alpha_1 + t_{1s} + \alpha_2 + t_{2s} - 2}{2(\alpha_1 + \alpha_2 + \alpha_3 + n - 3)}$. Denominamos o valor máximo desta densidade por $f^* = \sup_{\pi_1 \in (0, 1/2)} \{f_0(\pi_1 | \boldsymbol{t}_s)\} = f_0(\pi_1^* | \boldsymbol{t}_s)$.



(c) População Finita, n=80

(d) Superpopulação, n = 80

Figura 4.22: Valores da probabilidade a posteriori de $H_0: \pi_1 \leq \pi_2 \ (T_1 \leq T_2)$ para N = 100.

O segundo passo é o cálculo da evidência em favor de H_0 , para o qual é necessário definir a região tangente à hipótese nula

$$T_{t_s} = \{ \boldsymbol{\pi} \in S : f(\boldsymbol{\pi} | \boldsymbol{t_s}) > f^* \}, \text{ onde } S = \{ (\pi_1, \pi_2) \in \mathbb{R}^2_+ : \pi_1 + \pi_2 \leq 1 \},$$

e então calcular a evidência, dada por

$$ev(\{\pi_1 = \pi_2\}; \boldsymbol{t_s}) = 1 - \int_{T_{t_s}} \frac{1}{\beta(\boldsymbol{\alpha} + \boldsymbol{t_s})} \prod_{i=1}^3 \pi_i^{\alpha_i + t_{is} - 1} d\pi_1 d\pi_2$$

onde $\pi_3 = \pi_1 - \pi_2$. O cálculo da evidência de $\{\pi_1 = \pi_2\}$ é também feito através de simulação Monte Carlo. Neste procedimento, geramos uma amostra com 10⁴ pontos da distribuição Dirichlet acima e calculamos os valores da densidade a posteriori destes pontos. A evidência em favor de H_0 é então estimada pela proporção de elementos com densidade menor ou igual a f^* .

Sob a abordagem preditivista, seguimos os mesmos passos para o cálculo da evidência, mas de forma mais direta, pois o máximo sob H_0 é encontrado de forma simples e não é necessário utilizar simulações no cálculo da evidência. Sob a hipótese nula, o valor máximo f^* é a maior probabilidade a posteriori quando $T_1 = T_2$ e é dado por

$$f^* = max_{t \in \mathbb{N}} \left\{ P\left(T_1 = t, T_2 = t, T_3 = N - 2t \mid \sum_{i=1}^n \mathbf{Y}_i = \mathbf{t}_s\right) \right\} ,$$

que é bem simples de ser encontrado uma vez que já temos calculadas as probabilidades a posteriori. A evidência a favor de H_0 é então a soma de todas as probabilidades a posteriori que sejam inferiores ou iguais à f^* e pode ser escrita como

$$ev(\{T_1 = T_2\}; \boldsymbol{t_s}) = 1 - \sum_{\left\{\boldsymbol{t}: P(\boldsymbol{T} = \boldsymbol{t} | \sum_{i=1}^n \boldsymbol{Y_i} = \boldsymbol{t_s}) > f^*\right\}} P\left(\boldsymbol{T} = \boldsymbol{t} \left| \sum_{i=1}^n \boldsymbol{Y_i} = \boldsymbol{t_s} \right)\right\}.$$

A Figura 4.23 exibe as evidências em favor de H_0 em função da estatística suficiente (t_{1s}, t_{2s}) , para amostras de tamanho 20 e 80, considerando as abordagens preditivista para populações finitas e de superpopulação. Observamos, assim como no problema anterior, que para a amostra de tamanho 20, as superfícies formadas pelas evidências são muito parecidas,

enquanto que para n = 80 as diferenças são maiores, sendo uma superfície mais suave na Figura 4.23(d) (modelo de superpopulação) e as evidências crescem (e decrescem) mais rapidamente na Figura 4.23(c) (modelo de populações finitas).

Considere a amostra com n=20 (Figuras 4.23(a) e 4.23(b)) e suponha que o valor observado da estatística suficiente foi $t_s = (13, 4, 3)$. Então, sob a abordagem preditivista, a evidência em favor de H_0' obtida é 0,03, enquanto que, considerando o modelo de superpopulação, este valor sobe para 0,07. Note que, devido à permutabilidade e à uniformidade das distribuições a priori, os valores obtidos para as evidências em favor de H_0 seriam os mesmos se o resultado amostral fosse $t_s = (4, 13, 3)$.

Quando o tamanho da amostra aumenta em relação ao tamanho da população, as diferenças nos valores das evidências se destacam mais. Para uma amostra que produza $t_s = (33, 47, 0)$ ou $t_s = (47, 33, 0)$, a evidência calculada para populações finitas é 0,003, enquanto que sob o enfoque de superpopulação este valor sobe para 0,48. Para $t_s = (30, 40, 10)$ (ou $t_s = (40, 30, 10)$), estes valores são 0,02 e 0,48, respectivamente. Outras grandes diferenças também ocorrem, especialmente nas amostras em que os totais t_{1s} e t_{2s} são pequenos em relação ao valor t_{3s} .



Figura 4.23: Valores da evidência em favor de $H'_0: \pi_1 = \pi_2$ $(T_1 = T_2)$ para N = 100.

4.4.2 Testando proporções simultaneamente

Ainda considerando o modelo Multinomial, abordamos agora uma nova situação, na qual estamos interessados em testar conjuntamente todas as proporções (totais) do vetor π (**T**). Uma das possíveis aplicações de testes conjuntos/simultâneos é, por exemplo, num cenário de eleições majoritárias com 3 candidatos, como discutido em Fossaluza (2008). Neste cenário simplificado, podemos imaginar um pesquisador interessado em saber se haverá ou não segundo turno num determinado pleito. Do ponto de vista estatístico, avaliar a ocorrência ou não do segundo turno das eleições pode ser visto como um problema de decisão, no qual devemos optar por uma das hipóteses:

$$\begin{cases} H_0: \bigcap_{i=1}^3 \{\pi_i \le 0, 5\} \\ H_1: \bigcup_{i=1}^3 \{\pi_i > 0, 5\} \end{cases}.$$
(4.16)

Desse modo, H_0 indica que nenhum dos 3 candidatos terá mais do que 50% dos votos e, portanto, haverá segundo turno. Por outro lado, a hipótese alternativa H_1 indica que pelo menos um dos candidatos conseguirá mais do que a metade dos votos e, portanto, não será necessária a realização do segundo turno das eleições.

Para testarmos as hipóteses em (4.16), utilizamos o teste de Bayes baseado na probabilidade a posteriori da hipótese nula. Sob o modelo de superpopulação, calculamos a probabilidade de H_0 a partir das distribuições marginais dos elementos do vetor $\boldsymbol{\pi} = (\pi_1, \pi_2)$ (lembrando que $\pi_3 = 1 - \pi_1 - \pi_2$) da seguinte maneira:

$$\begin{split} P\left(H_{0}\left|\sum_{i=1}^{n}\boldsymbol{Y_{i}}=\boldsymbol{t_{s}}\right) &= P\left(\bigcap_{i=1}^{3}\{\pi_{i}\leq0,5\}\left|\sum_{i=1}^{n}\boldsymbol{Y_{i}}=\boldsymbol{t_{s}}\right)\right) \\ &= 1-P\left(\bigcup_{i=1}^{3}\{\pi_{i}>0,5\}\left|\sum_{i=1}^{n}\boldsymbol{Y_{i}}=\boldsymbol{t_{s}}\right)\right) \\ &= 1-\sum_{i=1}^{3}P\left(\pi_{i}>0,5\left|\sum_{i=1}^{n}\boldsymbol{Y_{i}}=\boldsymbol{t_{s}}\right)\right) \\ &= 1-\sum_{i=1}^{3}\int_{0.5}^{1}\frac{\Gamma\left(\sum_{k=1}^{3}\{\alpha_{k}+t_{ks}\}\right)}{\Gamma(\alpha_{i}+t_{is})\Gamma\left(\sum_{j\neq i}\{\alpha_{j}+t_{js}\}\right)}\pi_{i}^{\alpha_{i}+t_{is}-1}(1-\pi_{i})^{\sum_{j\neq i}\{\alpha_{j}+t_{js}\}-1}d\pi_{i} \; . \end{split}$$

Sob o enfoque preditivista para populações finitas, a hipótese nula de que não haverá segundo turno é $H_0: \bigcap_{i=1}^3 \{T_i \leq N/2\}$ e sua probabilidade é:

$$P\left(H_0\left|\sum_{i=1}^n \mathbf{Y_i} = \mathbf{t_s}\right) = \sum_{t \in D} P\left(T = t\left|\sum_{i=1}^n \mathbf{Y_i} = \mathbf{t_s}\right)\right|,$$

sendo $D = \left\{ (t_1, t_2, t_3) \in \mathbb{N}^3 : t_1 \leq \frac{N}{2}, t_2 \leq \frac{N}{2}, t_1 + t_2 \geq \frac{N}{2} e t_1 + t_2 + t_3 = N \right\}.$

A Figura 4.24 exibe as probabilidades da hipótese nula em função da estatística suficiente (t_{1s}, t_{2s}) , para N = 100. Assim como nos casos anteriores, observamos, para uma amostra de tamanho 80, uma superfície mais suave formada pelas probabilidades calculadas sob o modelo de superpopulação (Figura 4.24(d)). Os valores crescem (e decrescem) mais rapidamente sob o modelo de populações finitas (Figura 4.24(c)).

Considerando amostra de tamanho 20, as superfícies exibidas nas Figuras 4.24(a) e 4.24(b) são parecidas, mas ainda ocorrem diferenças entre as probabilidades calculadas pelos diferentes métodos. Por exemplo, se $t_s = (9, 2, 9)$, então a probabilidade de H_0 calculada sob a abordagem preditivista para populações finitas é 0,53, enquanto que sob o enfoque de superpopulação esse valor cai para 0,48. Os mesmos valores ocorrem quando os totais da amostra são $t_s = (2, 9, 9)$ ou $t_s = (9, 9, 2)$.

Para n = 80, as diferenças entre os métodos crescem, chegando a mais de 36%. Por exemplo, para $\mathbf{t}_s = (37, 5, 38)$ (ou para qualquer permutação dos valores deste vetor) a probabilidade de H_0 sob o modelo de superpopulação é 0,49, enquanto que para populações finitas este valor é 0,85. Se $\mathbf{t}_s = (38, 4, 38)$, as probabilidades são 0,79 para populações finitas e 0,42 considerando superpopulação.

Análises considerando uma população de tamanho 500 são apresentadas no Apêndice C.



(c) População Finita, n=80



Figura 4.24: Valores da probabilidade a posteriori de $H_0 : \bigcap_{i=1}^3 \{\pi_i \leq 0, 5\} \left(\bigcap_{i=1}^3 \{T_i \leq \frac{N}{2}\}\right)$ para N = 100.

54 DISTRIBUIÇÕES PREDITIVAS

Capítulo 5

Considerações Finais

Neste trabalho, estudamos a abordagem preditivista, introduzida por de Finetti e que é vista como uma forma menos "artificial"da abordagem bayesiana. Considerando uma população finita e a partir da suposição de permutabilidade, apresentamos as distribuições a posteriori dos parâmetros operacionais, obtidas sob a perspectiva preditivista, para alguns modelos usuais.

Os objetivos principais foram a construção de testes de hipóteses para os parâmetros operacionais utilizando as distribuições a posteriori construídas sob o enfoque preditivista para populações finitas e a comparação dos resultados destes testes com resultados obtidos a partir dos modelos bayesianos usuais de superpopulação. Foram consideradas hipóteses simples, compostas, precisas e não precisas e utilizados o teste de Bayes baseado na probabilidade a posteriori da hipótese nula e o FBST. Como no caso de hipóteses precisas (sob o modelo de superpopulação), a comparação entre densidades e probabilidades é de díficil interpretação, consideramos uma extensão do FBST para casos em que a distribuição a posteriori do parâmetro de interesse é discreta (modelo de populações finitas). A utilização deste procedimento mostrou-se adequada para a comparação dos resultados sob as duas abordagens: além das análises serem baseadas nas mesmas quantidades (probabilidades a posteriori de conjuntos tangentes à hipótese nula), vale ressaltar que no modelo de populações finitas, a quantidade de pontos com probabilidade a posteriori positiva é, em geral, muito grande, tornando a probabilidade da hipótese nula muito pequena.

Os primeiros modelos estudados foram o Bernoulli e o Poisson, nos quais o parâmetro

operacional de interesse é o total (média) populacional. Para estas situações, observamos que, dependendo da amostra obtida, os valores das probabilidades a posteriori da hipótese nula calculados sob a abordagem preditivista para populações finitas diferem dos calculados sob o enfoque de superpopulação. Além disso, esta diferença cresce conforme o tamanho amostral se aproxima do tamanho populacional. Quando utilizamos prioris informativas, observamos que as probabilidades das hipóteses calculadas sob o enfoque de superpopulação são mais influenciadas pelas distribuições a priori, sendo que as diferenças entre as probabilidades de H_0 quando passamos de uma priori uniforme para uma priori informativa são menores no modelo preditivista. No modelo Bernoulli, quando fixamos o total populacional e estamos interessados em hipóteses unilaterais da forma $H_0: \pi \leq \pi_0$ ($\theta \leq \theta_0$), os resultados dos testes de hipóteses obtidos a partir do modelo preditivista para populações finitas foram melhores do que os obtidos pelos modelos de superpopulação, considerando amostras grandes. Para amostras pequenas, as performances dos procedimentos são equivalentes.

Para o modelo Uniforme discreto, estudamos o máximo populacional. Através das análises realizadas, constatamos que as diferenças entre as probabilidades nas duas abordagens não estão ligadas somente à relação entre o tamanho amostral e o populacional, mas também são influenciadas pelo máximo populacional $X_{(N)}$, sendo que, quando este valor é muito menor que N, as variáveis podem assumir poucos valores em \mathbb{N} e, praticamente, não existem diferenças entre os métodos, mesmo com tamanhos amostrais grandes em relação ao tamanho populacional.

No caso multivariado, consideramos o modelo Multinomial de dimensão 3. Quando o objetivo é fazer alguma afirmação sobre a relação entre duas das três proporções envolvidas, digamos $\pi_1 \in \pi_2$, observamos que as maiores diferenças entre os métodos ocorrem quando, na amostra, a proporção de elementos da terceira categoria é muito maior que as proporções das outras duas categorias envolvidas em H_0 . Por exemplo, nos casos $H_0: \pi_1 \leq \pi_2 \in H'_0: \pi_1 = \pi_2$, as maiores diferenças ocorrem quando a proporção amostral $\hat{\pi}_3$ é próxima de 1. Isso ocorre, pois, sob a abordagem preditivista, os pontos que descrevem que todos (ou quase todos) os elementos da primeira e segunda categorias presentes na população já foram amostrados têm alta probabilidade. Quando o interesse é testar as três proporções simultaneamente, observamos novamente que as maiores diferenças ocorrem quando o valor de uma das 3

proporções observadas na amostra fica próximo de zero.

Em geral, comparando os modelos de populações finitas com os seus correspondentes de superpopulação, gerados a partir das mesmas suposições, observamos que ocorrem grandes diferenças nos resultados dos testes de hipótese quando o tamanho amostral é grande relativamente ao populacional. O crescimento e o decaimento das probabilidades e evidências da hipótese nula são mais suaves nos modelos de superpopulação e mais sensíveis às flutuações da estatística suficiente nos modelos preditivistas para populações finitas. Assim, consideramos que os modelos de populações finitas devem ser utilizados nas situações em que o tamanho da população é conhecido e o tamanho amostral é grande relativamente ao populacional.

5.1 Perspectivas Futuras

- Estudo das propriedades lógicas do FBST quando aplicado à distribuições discretas.
- Avaliação das diferenças entre as abordagens preditivista para populações finitas e de superpopulação para outras distribuições.
- Desenvolvimento de resultados analíticos para as diferenças entre os resultados dos testes de hipóteses produzidos pelas duas abordagens. Por exemplo, no caso Bernoulli, considerando as hipóteses:

$$\begin{cases} H_0: \theta \le \theta_0 & (\pi \le \pi_0) \\ H_1: \theta > \theta_0 & (\pi < \pi_0) \end{cases}$$

,

deve existir $n^* = n^*(N, \theta_0, \epsilon) \in \mathbb{N}$, com $N \in \mathbb{N}$, $\theta_0 \in \mathbb{N}$ e $\epsilon > 0$, tal que para $n > n^*$

$$\Big|P_{PF}\big(H_0\big|(x_1,\ldots,x_n)\big)-P_{Sup}\big(H_0\big|(x_1,\ldots,x_n)\big)\Big|>\epsilon,$$

para certos valores de $\sum_{i=1}^{n} X_i$, onde P_{PF} e P_{Sup} denotam probabilidades sob as abordagens preditivista para populações finitas e de superpopulação, respectivamente.

58 CONSIDERAÇÕES FINAIS

Apêndice A

Bernoulli



Figura A.1: Valores da probabilidade de $H_0: \pi \leq 0, 5$ $(\theta \leq \frac{N}{2})$ em função do total amostral para uma população de tamanho N = 1000.



Figura A.2: Valores da evidência em favor de H'_0 : $\pi = 0,5$ ($\theta = \frac{N}{2}$) em função do total amostral para uma população de tamanho N = 1000.



Figura A.3: Valores da probabilidade de $H_0: \pi \leq \pi_0$ ($\theta \leq N\pi_0$) em função do total amostral para uma população de tamanho N = 100 e n = 20.



Figura A.4: Valores da probabilidade de H_0 : $\pi \leq \pi_0$ ($\theta \leq N\pi_0$) em função do total amostral para uma população de tamanho N = 100 e n = 80.



Figura A.5: Valores da evidência em favor de H'_0 : $\pi = \pi_0$ ($\theta = N\pi_0$) em função do total amostral para uma população de tamanho N = 100 e n = 20.



Figura A.6: Valores da evidência em favor de H'_0 : $\pi = \pi_0$ ($\theta = N\pi_0$) em função do total amostral para uma população de tamanho N = 100 e n = 80.

62 APÊNDICE A
Apêndice B

Poisson



Figura B.1: Valores da probabilidade de H_0 : $\lambda \leq 10$ ($\Lambda \leq 10N$) em função da média amostral.



Figura B.2: Valores da evidência em favor de H_0 : $\lambda = 10$ ($\Lambda = 10N$) em função da média amostral.

Apêndice C

Multinomial



(c) População Finita, n = 400 (d) Superpopulação, n = 400

Figura C.1: Valores da probabilidade a posteriori de H_0 : $\pi_1 \leq \pi_2$ $(T_1 \leq T_2)$ para N = 500.



Figura C.2: Valores da evidência em favor de H'_0 : $\pi_1 = \pi_2$ $(T_1 = T_2)$ para N = 500.



Figura C.3: Valores da probabilidade a posteriori de $H_0 : \bigcap_{i=1}^3 \{\pi_i \leq 0, 5\} \left(\bigcap_{i=1}^3 \{T_i \leq \frac{N}{2}\}\right)$ para N = 500.

68 APÊNDICE C

Referências Bibliográficas

- **Bonassi (2009)** F. V. Bonassi. Permutabilidade de quantidades aleatórias e a falácia do apostador. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 5
- **DeGroot (1970)** M. H. DeGroot. *Optimal statistical decisions*. New York: McGraw-Hill. Citado na pág. 11
- **Diaconis (1977)** P. Diaconis. Finite forms of de finetti's theorem on exchangeability. Synthese, 36(2):271-281. Citado na pág. 8
- Fortini e Petrone (2012) S. Fortini e S. Petrone. Predictive construction in bayesian nonparametrics. *Brazilian Journal of Probability and Statistics*, 3. Citado na pág. 1
- **Fossaluza (2008)** V. Fossaluza. Testes de hipóteses em eleições majoritárias. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 10, 51
- Heath e Sudderth (1976) D. Heath e W. Sudderth. De finetti's theorem in exchangeable sequences. The American Statistician, 30:188–189. Citado na pág. 7, 8
- Iglesias (1993) P. L. Iglesias. Formas finitas do teorema de De Finetti: a visão preditivista da inferência estatística em populações finitas. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 2, 6, 7, 8, 17, 42, 43
- Iglesias et al. (2009) P.L. Iglesias, R.H. Loschi, C.A.B. Pereira e S. Wechsler. A note on extendibility and predictivistic inference in finite populations. *Brazilian Journal of Probability and Statistics*, 23(2):216-226. Citado na pág. 2, 17
- Izbicki (2010) R. Izbicki. Classes de testes de hipóteses. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 12
- Lindley e Phillips (1976) D. V. Lindley e L. D. Phillips. Inference for a bernoulli process. The American Statistician, 30:112–119. Citado na pág. 6
- Madruga e Pereira (2005) M. Madruga e C. A. B. Pereira. Power of fbst: Standard examples. Instituto Interamericano de Estadística, Estadística, 57:1-9. Citado na pág. 12
- Madruga et al. (2001) M.R. Madruga, L.G. Esteves e S. Wechsler. On the bayesianity of pereira-stern tests. *Test*, 10(2):291-299. Citado na pág. 11, 12
- Mendel (1994) M. B. Mendel. Operational parameters in bayesian models. *Test*, 3:195–206. Citado na pág. 2

- **Pereira e Stern (1999)** C. A. B. Pereira e J. M. Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy Journal*, 1:104–115. Citado na pág. 11, 12
- Pereira et al. (2008) C. A. B. Pereira, J. M. Stern e S. Wechsler. Can a significance test be genuinely bayesian? *Bayesian Analysis*, 3:79–100. Citado na pág. 9, 12

Schervish (1995) M. J. Schervish. Theory of Statistics. Springer. Citado na pág. 1, 7, 8, 13

Wechsler (1993) S. Wechsler. Exchangeability and predictivism. *Erkenntnis*, 38:343-350. Citado na pág. 1, 8, 18, 29