Testes de Significância Bayesianos.
Aplicações a Distribuições Discretas
André Luiz Silva Samartini

Dissertação Apresentada ao Instituto de Matemática
e Estatística da Universidade de São Paulo
para Obtenção do Grau de Mestre em Estatística

Área de Concentração: Estatística

Orientador: Prof. Dr. Carlos Alberto de Bragança Pereira

Durante o curso de Mestrado e a elaboração do trabalho, o autor recebeu apoio financeiro
do CNPq.

- São Paulo, agosto de 2000 -

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por André Luiz Silva Samartini
e aprovada pela comissão julgadora.

São Paulo, 10 de agosto de 2000.

Banca examinadora:

- Prof. Dr. Carlos Alberto de Bragança Pereira (Orientador) - IME - USP
- Prof. Dr. Julio da Motta Singer - IME - USP
- Prof. Dr. Eduardo Massad - Faculdade de Medicina - USP

Agradecimentos

Em primeiro lugar eu gostaria de agradecer à minha família, que me apoiou desde o jardim de infância até o final do mestrado. Eles sempre valorizaram o estudo, as vezes até demais. Mas agora eu consigo olhar para trás e ver o quanto eu conquistei e o quanto eles me ajudaram.

Também quero agradecer aos meus vários amigos da faculdade (incluindo alguns professores), pois nesses vários anos de faculdade o IME foi a minha segunda casa. Na minha memória não ficarão apenas teorias estatísticas, mas também muitos momentos que passei durante a graduação e o mestrado.

Por fim, agradeço ao meu orientador, o "Carlinhos", pois graças a ele em nenhum momento me senti inseguro do que estava fazendo e sempre me senti motivado a melhorar o meu trabalho.

Abstract

The p-value, used as a tool to test hypothesis, is part of the scientific language since it is a measure of the evidence of a particular hypotheses. In many cases, the calculus of the p-value does not take into account the alternative hypotheses, only the null hypotheses. In this work it is studied a measure of evidence called Bayesian P-value that considers both hypothesis in its calculus and uses opinions represented by probability distributions *a priori* in the parametric space. We treat the case of precise hypothesis, which are defined by sets in the parametric space whose dimension is smaller than the original space. Surface integrals are used in the calculus of the Bayesian P-value. The examples presented are restricted to discrete distributions such as contingency tables and comparison of two Poisson distributions.

Resumo

O p-value, usado como ferramenta para testar hipóteses, já é parte da linguagem científica pois acredita-se que o mesmo é uma medida da evidência da validade de uma hipótese específica. Muitas vezes o cálculo do p-value - também chamado de nível descritivo - não leva em consideração a hipótese alternativa, mas apenas a hipótese nula. Discute-se neste trabalho uma medida de evidência, denominada P-value Bayesiano, que além de incorporar em seu cálculo ambas as hipóteses, leva em consideração opiniões representadas por distribuições de probabilidades *a priori* no espaço paramétrico. Tratamos do caso de hipóteses nulas precisas, que são definidas por conjuntos do espaço paramétrico cuja dimensão é menor do que o espaço original. Integrais de superfície são usadas no cálculo do P-value Bayesiano. Os exemplos apresentados se restringem a distribuições discretas, mais especificamente a tabelas de contingência e a comparação de distribuições de Poisson.

Sumário

1	Introdução	1
2	Definição do P-value Bayesiano	3
2.1	Teste de Neyman-Pearson para duas hipóteses simples	3
2.2	Hipóteses compostas	5
2.3	Cálculo do P-value	7
3	Aplicações	8
3.1	Comparação de Binomiais (teste de homogeneidade)	8
3.1.1	Definição da hipótese	8
3.1.2	O teste clássico	9
3.1.3	O teste de Bayes	10
3.2	Comparação de duas Poisson independentes.	12

3.2.1	O teste clássico	12
3.2.2	O teste de Bayes	13
3.3	Teste de independência em tabelas de contingência 2x2	15
3.3.1	O Teste clássico	15
3.3.2	O teste de Bayes	16
3.4	Teste de equilíbrio populacional	18
3.4.1	Definição de equilíbrio populacional	18
3.4.2	O teste clássico	19
3.4.3	O teste de Bayes	20
4	Comparação entre as medidas de evidência	23
4.1	Exemplos de inconsistência no valor do p-value clássico	23
4.2	Comparação entre as evidências obtidas pelos dois métodos	25
4.2.1	Teste de binomiais	25
4.2.2	Comparação de duas Poisson	27
4.2.3	Teste de independência	28
4.2.4	O teste de equilíbrio de Hardy-Weinberg	29
4.3	Comparação entre os erros reportados pelos dois métodos	30

4.3.1	Teste de binomiais	31
4.3.2	Teste de Poisson	32
4.3.3	Teste de independência	33
4.3.4	Teste de Hardy-Weinberg	35
A	Aspectos Computacionais	38
A.1	Simulação	38
A.2	Cálculo da verossimilhança	40
A.3	Tempo de computação	41
B	Gráficos	42

Capítulo 1

Introdução

Este trabalho apresenta o cálculo do P-value Bayesiano sugerido em Pereira & Wechsler (1993) e usa a técnica de integrais de superfície apresentada por Irony & Pereira (1995) no caso de hipóteses compostas. Como são utilizadas distribuições de probabilidades no espaço paramétrico, pode-se dizer que o trabalho está dentro do paradigma Bayesiano, embora interpretações sob o ponto de vista da estatística clássica também sejam possíveis se admitirmos um sistema de preferências sobre o conjunto dos valores possíveis do parâmetro, o espaço paramétrico.

Ao contrário dos diversos testes clássicos, o teste de significância apresentado é exato pois não utiliza aproximações assintóticas. Já os testes exatos padrões utilizam funções de verossimilhança parciais, normalmente condicionais a estatísticas suficientes específicas (Basu (1977) ou Cox (1958)). Como exemplos mais conhecidos, citamos os testes de Fisher para tabelas de contingência, o teste de equilíbrio de Hardy-Weinberg (EHW) e o teste de comparação de distribuições de Poisson.

A intuição por trás do cálculo do p-value é a ordenação do espaço amostral. Com o espaço amostral ordenado calcula-se o p-value, que nada mais é do que a probabilidade do conjunto dos pontos amostrais mais extremos do que o ponto amostral efetivamente observa-

do. Essa probabilidade é condicional à hipótese nula ser verdadeira. O problema que surge nesse procedimento é a ordenação do espaço amostral. Nos exemplos apresentados por Pereira & Wechsler (1993) os pontos extremos, definidos sem levar em consideração as hipóteses alternativas, são mais extremos sob a hipótese alternativa do que sob a nula. Assim, os pontos extremos usados no cálculo do p-value podem levar à rejeição da hipótese nula e aceitação da alternativa mesmo quando a probabilidade desses pontos sob a hipótese alternativa é ainda menor. Como o cálculo do P-value é baseado na comparação das verossimilhanças sob as duas hipóteses, esse problema é eliminado.

Deve-se ressaltar que no cálculo usual da razão de verossimilhanças, utiliza-se o supremo da verossimilhança nos subconjuntos do espaço paramétrico gerados pelas hipóteses. No teste apresentado nesta dissertação, será calculada uma razão de verossimilhanças ponderadas, em cada um dos subconjuntos gerados pelas hipóteses. Essa ponderação é feita através de distribuições a priori definidas nos espaços paramétricos associados a cada hipótese. Uma consequência importante é que, ao comparar as verossimilhanças ponderadas, o teste minimiza a soma das probabilidades dos erros do tipo I e do tipo II.

O objetivo da dissertação é, além de revisar o trabalho feito em Pereira & Wechsler (1993), desenvolver programas para realizar o cálculo do P-value Bayesiano no software MATLAB. Os aspectos computacionais são discutidos no apêndice.

Capítulo 2

Definição do P-value Bayesiano

Neste capítulo será construído um teste de hipóteses utilizando o Lema de Neyman-Pearson e calcular-se-á o P-value Bayesiano a partir deste teste.

2.1 Teste de Neyman-Pearson para duas hipóteses simples

Seja $f(d|\omega)$ a função densidade de probabilidade, onde d são os dados obtidos de uma amostra e ω o parâmetro de interesse. O parâmetro θ será o indicador de hipóteses, ou seja, $\theta = i, i = 0, 1$ se $\omega \in \Omega_i$, onde $\Omega_0 \cup \Omega_1 = \Omega$ e $\Omega_0 \cap \Omega_1 = \emptyset$. Ω_0 e Ω_1 são disjuntos. Suponha que queiramos testar $H_0: \theta = 0$ vs $H_1: \theta = 1$.

Um teste de hipótese é uma função binária δ dos dados tal que se $\delta(d) = 1$, rejeita-se H_0 em favor de H_1 e se $\delta(d) = 0$ não se rejeita H_0 em prejuízo de H_1 . Desse modo, há dois tipos de erros que podem ser cometidos: o erro do tipo I, que ocorre quando se rejeita H_0 quando esta é verdadeira e o erro do tipo II, que ocorre quando não se rejeita H_0 quando esta é falsa. As probabilidades de erro do tipo I e do tipo II serão representadas, respectivamente,

por $\alpha(d)$ e $\beta(d)$. O critério de seleção de um teste é estabelecido pelo seguinte teorema:

Teorema 1: Sejam duas constantes a e b e sejam também $f_0(d)$ e $f_1(d)$ as funções densidade de probabilidade sob H_0 e H_1 , respectivamente. Seja δ^* um teste definido como a seguir:

$$\delta^*(d) = 0 \quad \text{se} \quad af_0(d) \geq bf_1(d)$$

$$\text{e } \delta^*(d) = 1 \quad \text{se} \quad af_0(d) < bf_1(d)$$

Então, $a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta)$ (DeGroot 1989) para qualquer teste δ .

Os valores de a e b determinam o grau de importância dos erros do tipo I e II, respectivamente. Se $\frac{b}{a} = 1$, diz-se que os dois tipos de erro são igualmente importantes, ao passo que se $\frac{b}{a} < 1 (> 1)$ a probabilidade de erro do tipo I (tipo II) é a que deve ser mais minimizada.

Teorema 2: O teste δ^* definido anteriormente é uma solução Bayesiana para o problema do teste de H_0 contra H_1 .

Demonstração: Seja $L(\theta, \delta)$ a função de perda para o teste δ . No corpo da tabela a seguir são mostradas as perdas para cada valor de δ e θ :

	$\delta = 0$	$\delta = 1$
$\theta = 0$	0	l_0
$\theta = 1$	l_1	0

Vamos supor que as preferências a priori pelas hipóteses sejam dadas por $P(\theta = i) = \xi_i, i = 0, 1$. Assim, a perda esperada $r(\delta)$ do teste δ será

$$r(\delta) = \xi_0 E\{L(\theta, \delta) | \theta = 0\} + \xi_1 E\{L(\theta, \delta) | \theta = 1\} = \xi_0 l_0 \alpha(\delta) + \xi_1 l_1 \beta(\delta).$$

Note que δ é um teste que minimiza a perda esperada. Basta fazer $a = \xi_0 l_0$ e $b = \xi_1 l_1$. Pelo Teorema 1, qualquer que seja outro teste δ' , $r(\delta) < r(\delta')$.

Portanto, δ é um teste de Bayes, pois minimiza a perda esperada.

Até agora, foi tratado o caso em que as hipóteses a serem testadas são simples. A seguir, será apresentada uma técnica para construção de testes com hipóteses compostas.

2.2 Hipóteses compostas

No caso de hipóteses simples, podemos aceitar ou rejeitar uma hipótese comparando as verossimilhanças sob a hipótese nula e sob a hipótese alternativa. Se considerarmos a mesma perda para os dois tipos de erros, basta verificar sob qual hipótese a verossimilhança é maior.

Suponha agora que queiramos testar $H_0 : \omega \in \Omega_0$ e $H_1 : \omega \in \Omega_1$. Como definir a verossimilhança sob as hipóteses? Pode-se, por exemplo, calcular a verossimilhança média sob H_0 (H_1), dando peso igual para todos os pontos de Ω_0 (Ω_1). Do ponto de vista Bayesiano, seria equivalente a considerar uma distribuição a priori sob cada hipótese para o parâmetro em questão. Utilizando integrais de superfície, pode-se calcular uma verossimilhança ponderada, de acordo com a distribuição a priori definida. Uma consequência importante é que, ao comparar as verossimilhanças ponderadas, o teste minimiza a soma das probabilidades dos erros do tipo I e II. A seguir, será descrito formalmente o teste proposto em Pereira (1985), que segue essa linha de raciocínio.

As hipóteses serão definidas em termos de uma partição do espaço amostral, isto é, $H_0 : \omega \in \Omega_0$ e $H_1 : \omega \in \Omega_1$.

A distribuição a priori P de ω sob H_i ($i = 0, 1$) está associada a uma função densidade de probabilidade definida por:

$$g(\omega|\omega \in \Omega_i) = g_i(\omega) = \frac{g(\omega)I_i(\omega)}{\int g(\omega)d\Omega_i},$$

em que $I_i(\omega) = 1$ se $\omega \in H_i$ e 0 caso contrário. $\int g(\omega)d\Omega_i$ é a integral de linha (em $\mathbb{R}^2$) ou a integral de superfície (em $\mathbb{R}^n, n > 2$) no subconjunto Ω_i . Deve-se notar que $g_i(\omega)$

é a densidade condicional de ω dado que $\omega \in \Omega_i$ (desde que Ω_i não tenha probabilidade zero na medida P , pois caso contrário o cálculo da distribuição condicional não estaria bem definido).

Vamos denotar a verossimilhança associada ao modelo por $L(\omega|d)$. A função de verossimilhança nada mais é do que o valor da função densidade de probabilidade amostral $f(\cdot|\omega)$ calculada nos dados observados d .

Com todas as definições anteriores e utilizando a fórmula de Bayes, pode-se obter a distribuição a posteriori de ω :

$$g(\omega|d) = \frac{L(\omega|d)g(\omega)}{\int_{\Omega} L(\omega|d)g(\omega)d\omega}.$$

Da mesma maneira, pode-se definir a distribuição a posteriori de ω sob H_i , que é dada por

$$g(\omega|d, \omega \in \Omega_i) = g_i(\omega|d) = \frac{L(\omega|d)g(\omega)I_i(\omega)}{\int_{\Omega} L(\omega|d)g(\omega)d\Omega_i}.$$

Note que o denominador da fração é a integral de linha (ou superfície) no subconjunto Ω_i .

A função preditiva (ou densidade preditiva) dos dados é definida por:

$$f(d) = \int_{\Omega} f(d|\omega)g(\omega)d\omega = \int_{\Omega} L(\omega|d)g(\omega)d\omega.$$

Conseqüentemente, a função preditiva sob H_i pode ser escrita por

$$f_i(d) = \frac{\int_{\Omega_i} L(\omega|d)g(\omega)d\Omega_i}{\int_{\Omega_i} g(\omega)d\Omega_i}.$$

Podemos notar que $f_i(d)$ é a probabilidade condicional de d dado $\omega \in \Omega_i$. Portanto, a probabilidade a posteriori do evento $\{\omega \in \Omega_i\}$ pode ser obtida através de $f_i(d)$, utilizando a fórmula de Bayes, ou seja:

$$\xi_i(d) = \frac{\xi_i f_i(d)}{\sum \xi_i f_i(d)}.$$

Antes de se observar os dados, o analista dever ter uma priori para a probabilidade de suas hipóteses serem verdadeiras (ξ_0 e ξ_1). Para decidir em favor de H_0 ou H_1 , deve-se comparar esses dois valores. Essa comparação é feita através da razão (odds) a priori em favor de H_0 , $R_{01} = \frac{\xi_0}{\xi_1}$. Um teste bayesiano será favorável a H_0 se $R_{01} > c$, onde c é uma constante que depende do grau de importância dos erros do tipo I e do tipo II. De modo análogo, pode ser definida a razão em favor de H_1 .

Uma vez observados os dados, o conhecimento do analista é então atualizado. A razão em favor de H_0 a posteriori é então $R_{01}(d) = \frac{\xi_0(d)}{\xi_1(d)}$. Da mesma maneira, o teste (bayesiano) generalizado de Neyman-Pearson será favorável a H_0 se $R_{01} > c$.

2.3 Cálculo do P-value

Para calcular o P-value, deve-se fazer uma ordenação dessa razão de verossimilhanças de acordo com os pontos amostrais; portanto, a constante c aqui não é relevante. Um ponto amostral d_1 é mais extremo que um ponto amostral d_2 se $R_{01}(d_1) \leq R_{01}(d_2)$. O P-value associado ao ponto d é o número

$$p = \int I(r) f_0(r) dr.$$

em que $I(r)$ é o conjunto $\{r : R_{01}(r) < R_{01}(d)\}$

Em outras palavras, o P-value é a soma dos valores da função densidade de probabilidade sob H_0 , soma esta feita no conjunto de pontos amostrais onde a razão de verossimilhança é menor que a razão de verossimilhanças observada no ponto d .

Capítulo 3

Aplicações

3.1 Comparação de Binomiais (teste de homogeneidade)

3.1.1 Definição da hipótese

Suponha que x e y são observações de duas variáveis binomiais independentes X e Y , cujos parâmetros são $(m;p)$ e $(n;q)$. A seguinte tabela representa as observações e as probabilidades (entre parêntesis) associadas ao problema:

	Sucesso	Fracasso	Total
População 1	$x(p)$	$m - x(1 - p)$	m
População 2	$y(q)$	$n - y(1 - q)$	n
Total	$x + y$	$m + n - x - y$	$m + n$

Deseja-se testar as seguintes hipóteses:

$$H_0 : p = q$$

contra a alternativa

$$H_1 : p \neq q$$

3.1.2 O teste clássico

Os testes clássicos mais usados são o qui-quadrado e o teste exato de Fisher, descritos a seguir.

3.1.2.1 - O teste qui-quadrado

Sejam o_{ij} o valor observado e $e_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$ o valor esperado de cada casela. Sob H_0 , a estatística

$\chi^2 = \sum \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$ tem distribuição qui-quadrado com 1 grau de liberdade. O p-value é a área da distribuição à direita do valor observado.

Deve-se ressaltar que, como essa distribuição é assintótica, a aproximação pode não ser boa para amostras pequenas.

3.1.2.2 - O teste exato de Fisher

Sob H_0 , podemos obter a distribuição exata da frequência das caselas, condicionando na frequência de ambas as marginais. A distribuição da frequência das caselas obtida é a distribuição hipergeométrica, expressa abaixo:

$$\frac{\binom{m}{x} \binom{n}{y}}{\binom{m+n}{x+y}}$$

Dados os totais marginais, o valor de x determina os valores das outras caselas.

O p-value (unicaudal) calculado pelo teste de Fisher é a soma das probabilidades da distribuição acima para valores de x tão favoráveis à hipótese alternativa quanto o valor de

x observado. Por exemplo:

Se a tabela observada for

3	2	5
1	4	5
4	6	10

o p-value será dado pela soma de probabilidades das seguintes tabelas:

3	2	5	4	1	5
1	4	5	0	5	5
4	6	10	4	6	10

Para calcular o p-value bicaudal, somam-se as probabilidades de todas as tabelas (condicionais) com probabilidade menor ou igual à observada. No exemplo anterior, seriam somadas as probabilidades das seguintes tabelas:

3	2	5	1	4	5	0	5	5	4	1	5
1	4	5	3	2	5	4	1	5	0	5	5
4	6	10	4	6	10	4	6	10	4	6	10

3.1.3 O teste de Bayes

As hipóteses nesse problema definem a seguinte partição do espaço paramétrico:

$$\Omega_0 = \{(p, q) : 0 \leq p \leq 1, 0 \leq q \leq 1 \text{ e } p = q\}$$

$$\Omega_1 = \{(p, q) : 0 \leq p \leq 1, 0 \leq q \leq 1 \text{ e } p \neq q\}$$

A adoção de distribuições na classe conjugada torna o cálculo mais simples, pois a distribuição a posteriori é a mesma a menos dos parâmetros, que são atualizados. Pode-se também usar outras distribuições que não pertencem à classe conjugada da distribuição

binomial, caso haja interesse. Neste trabalho, o sistema de preferência sob Ω será dado pelo produto de duas distribuições Beta independentes, pois esta pertence à classe conjugada da distribuição binomial. Simbolicamente,

$$p \sim Be(a, b) \text{ e } q \sim Be(c, d).$$

A função $g(\omega)$ pode então ser escrita como

$$g(\omega) = g(p, q) = \frac{\Gamma(a+b)\Gamma(c+d)}{\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(d)} p^{a-1} (1-p)^{b-1} q^{c-1} (1-q)^{d-1},$$

em que $\Gamma(\cdot)$ é a função gama, dada por:

$$\Gamma(a) = \int x^{a-1} e^{-x} dx, \text{ para } a > 0. \text{ Se } a \text{ for inteiro, } \Gamma(a+1) = a!$$

A verossimilhança é expressa por

$$L(\omega|d) = \binom{m}{x} \binom{n}{y} p^x (1-p)^{m-x} q^y (1-q)^{n-y}.$$

A distribuição a posteriori também será produto de duas Betas independentes, com os seguintes parâmetros:

$$p \sim Be(A, B) \text{ e } q \sim Be(C, D),$$

$$\text{em que } A = a + x, B = b + m - x, C = c + y \text{ e } D = d + n - y$$

A função preditiva dos dados sob H_0 é dada por

$$f_0(x, y) = \binom{m}{x} \binom{n}{y} \frac{B(A+C-1, B+D-1)}{B(a+c-1, b+d-1)}$$

$$\text{em que } B(a, b) \text{ é a função beta no ponto } (a, b), \text{ ou seja, } B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Sob H_1 , tem-se:

$$f_1(x, y) = \binom{m}{x} \binom{n}{y} \frac{B(A, B)B(C, D)}{B(a, b)B(c, d)}$$

Considerando $\xi_0 = \xi_1 = \frac{1}{2}$, a razão de verossimilhanças em favor de H_0 é:

$$R_{01}(x, y) = \frac{B(A+C-1, B+D-1)B(a, b)B(c, d)}{B(a+c-1, b+d-1)B(A, B)B(C, D)}$$

Essa equação pode ser simplificada, se considerarmos uma priori uniforme:

$$R_{01}(x, y) = \frac{\binom{m}{x}\binom{n}{y}}{\binom{m+n}{x+y}} \frac{(m+1)(n+1)}{(m+n+1)}$$

Deve-se notar que o primeiro termo é o termo hipergeométrico, usado no teste exato de Fisher.

3.2 Comparação de duas Poisson independentes.

Suponha que $x \sim \text{Poisson}(m\lambda)$ e $y \sim \text{Poisson}(n\mu)$, sendo que x e y são números inteiros não negativos e $\lambda, \mu \in \mathfrak{R}$. A distribuição de probabilidade é dada por:

$$p(x|\lambda) = \frac{e^{-m\lambda}(m\lambda)^x}{x!} \text{ e } p(y|\mu) = \frac{e^{-n\mu}(n\mu)^y}{y!}$$

As hipóteses a serem testadas são

$$H_0 : \lambda = \mu$$

vs.

$$H_1 : \lambda \neq \mu$$

3.2.1 O teste clássico

O teste clássico utilizado para testar a igualdade de duas Poisson é feito através da probabilidade de observar o valor x dado $x + y$. Para facilitar, fixaremos $m = n$.

Temos então

$$P(X = x, Y = y) = P(X = x | X + Y = s)P(X + Y = s)$$

A expressão a seguir mostra que $X | X + Y$ tem distribuição binomial com parâmetro $\frac{\lambda}{\lambda + \mu}$:

$$P(X = x | X + Y = s) = \frac{e^{-n\lambda}(n\lambda)^x}{x!} \frac{e^{-n\mu}(n\mu)^{s-x}}{(s-x)!} \frac{s!}{e^{-n(\mu+\lambda)}(n\lambda+n\mu)^s} = \frac{s!}{x!(s-x)!} \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(\frac{\mu}{\lambda+\mu}\right)^{s-x}$$

Pode-se observar que testar $\lambda = \mu$ é o mesmo que testar $\frac{\lambda}{\lambda + \mu} = \frac{1}{2}$.

Como $\frac{\lambda}{\lambda + \mu}$ tem variação independente de $\lambda + \mu$, podemos inferir sobre $\frac{\lambda}{\lambda + \mu}$ utilizando apenas o termo $P(X = x | X + Y = s)$ através da estatística $\frac{X}{X+Y}$. Essa estatística é chamada de "Corte de Bardorff-Nielsen".

E $X | X + Y$ tem distribuição Binomial($X + Y, 0.5$) sob H_0 .

Com esse resultado, calcula-se o p-value da forma usual.

3.2.2 O teste de Bayes

As hipóteses definem a seguinte partição do espaço paramétrico:

$$\Omega_0 = \{(\lambda, \mu) : \lambda > 0, \mu > 0 \text{ e } \lambda = \mu\}$$

$$\Omega_1 = \{(\lambda, \mu) : \lambda > 0, \mu > 0 \text{ e } \lambda \neq \mu\}$$

A verossimilhança para esse problema é representada por:

$$L(\omega | x, y) = \frac{(m\lambda)^x}{x!} \frac{(n\mu)^y}{y!} e^{-m\lambda} e^{-n\mu}$$

Neste caso, será adotada a distribuição conjunta de duas Gammas independentes para representar a priori para $\omega = (\lambda, \mu)$, pois esta distribuição faz parte da classe de distribuições

conjugadas da Poisson. Portanto,

$\lambda \sim G(a, b)$ e $\mu \sim G(c, d)$, o que implica as seguintes distribuições a posteriori:

$$\lambda|x \sim G(a+x, b+m) \text{ e } \mu \sim G(c+y, d+n)$$

A distribuição conjunta a posteriori é dada por:

$$g(\omega|d) = g(\lambda, \mu|d) = \frac{(b+m)^{(a+x)}}{\Gamma(a+x)} \lambda^{a+x-1} e^{-(b+m)\lambda} \frac{(d+n)^{(c+y)}}{\Gamma(c+y)} \mu^{c+y-1} e^{-(d+n)\mu}$$

De posse dessas equações, pode-se calcular a função preditiva sob H_0 , que é escrita como:

$$f_0(x, y) = \frac{m^x n^y}{x!y!} \frac{\Gamma(A+C-1)(b+d)^{a+c-1}}{(B+D)^{A+C-1} \Gamma(a+c-1)}$$

em que: $A = a + x, B = b + m, C = c + y, D = d + n$

Sob H_1 , a função preditiva dos dados é dada por

$$f_1(x, y) = \frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \frac{m^x n^y}{x!y!} \frac{\Gamma(A)\Gamma(C)}{B^A D^C}$$

Fazendo $\xi_0 = \xi_1 = \frac{1}{2}$, a razão de verossimilhanças em favor de H_0 é:

$$R_{01}(x, y) = \frac{\Gamma(A+C-1)\Gamma(a)\Gamma(c)}{\Gamma(A)\Gamma(C)\Gamma(a+c-1)} \frac{B^A D^C}{(B+D)^{A+C-1}} \frac{(b+d)^{a+c-1}}{b^a d^c}$$

Considerando duas exponenciais independentes com médias iguais a 1 como priori (isto é, $a = b = c = d = 1$) e $m = n$, as expressões são simplificadas:

$$f_0(x, y) = \frac{n^{x+y}}{x!y!} \frac{(x+y)!2}{(2n+2)^{x+y+1}} = \binom{x+y}{x} \left(\frac{n}{n+1}\right)^{x+y} \left(\frac{1}{2}\right)^{x+y} \frac{1}{n+1}$$

$$f_1(x, y) = \frac{n^{x+y}}{x!y!} \frac{x!y!}{(n+1)^{(x+1)}(n+1)^{(y+1)}} = \frac{n^{x+y}}{(n+1)^{(x+y+2)}} = \left(\frac{n}{n+1}\right)^{x+y} \left(\frac{1}{n+1}\right)^2$$

e portanto

$$R_{01}(x, y) = \binom{x+y}{x} \left(\frac{1}{2}\right)^{x+y} (n+1)$$

3.3 Teste de independência em tabelas de contingência 2x2

As tabelas de contingência aqui consideradas são geradas por indivíduos de uma amostra que são classificados em uma das categorias combinadas existentes. Cada indivíduo deve ser classificado em uma e só uma categoria combinada. Em outras palavras, as categorias dessa classificação são exaustivas e mutuamente exclusivas. A notação utilizada para a frequência observada em cada casela está indicada na tabela a seguir:

	Classif.1	Classif.2	Total
Classif.1	n_{11}	n_{12}	$n_{1\bullet}$
Classif.2	n_{21}	n_{22}	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

Os parâmetros de interesse são:

	Classif.1	Classif.2	Total
Classif.1	p_{11}	p_{12}	p
Classif.2	p_{21}	p_{22}	$(1-p)$
Total	q	$1-q$	1

3.3.1 O Teste clássico

O teste clássico mais utilizado é o qui-quadrado, já apresentado no item 3.1.2.1.

A estatística obtida tem 1 grau de liberdade. O p-value é a área sob a distribuição para valores maiores que o observado.

3.3.2 O teste de Bayes

A distribuição considerada nesse caso é a multinomial. A escolha para a distribuição a priori será a Dirichlet com parâmetros a_{11} , a_{21} , a_{12} e a_{22} , que é conjugada para a distribuição multinomial. A função densidade da Dirichlet é proporcional a

$$p_{11}^{a_{11}-1} p_{12}^{a_{12}-1} p_{21}^{a_{21}-1} p_{22}^{a_{22}-1}$$

A função de verossimilhança é proporcional a

$$L(\omega|d) = L(p_{11}, p_{12}, p_{21}, p_{22} | n_{11}, n_{12}, n_{21}, n_{22}) = p_{11}^{n_{11}} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}}$$

Os parâmetros A_{ij} da distribuição a posteriori, que também é Dirichlet, são obtidos somando-se a frequência de cada casela com o parâmetro a priori correspondente, isto é, $A_{ij} = a_{ij} + n_{ij}$, para $i, j = 1, 2$.

A partição do espaço paramétrico que representa a hipótese de independência pode ser escrita como

$$H_0 : \omega = (p_{11}, p_{12}, p_{21}, p_{22}) \in [0, 1]^4 \text{ com } w_1 = pq, w_2 = p(1 - q), w_3 = (1 - p)q, w_4 = (1 - p)(1 - q), 0 \leq p, q \leq 1$$

e a hipótese alternativa,

$$H_1 : \omega \in \{(p_{11}, p_{12}, p_{21}, p_{22}) : p_{ij} > 0, \sum_{i,j} p_{ij} = 1; i, j = 1, 2\}$$

Para calcular a integral de superfície associada a H_0 , será utilizado o método descrito por Courant & John (1974).

Deve-se achar o fator Δ usado na integral de superfície. O primeiro passo para encontrar esse fator é calcular a matriz de derivadas de $(p_{11}, p_{12}, p_{21}, p_{22})$ em relação a p e q , que é dada por:

$$M = \begin{bmatrix} q & p \\ (1-q) & -p \\ -q & (1-p) \\ -(1-q) & -(1-p) \end{bmatrix}$$

Os seis determinantes das matrizes quadradas de posto 2 formadas pelas linhas de M são:

$$D_1 = -p, D_2 = q, D_3 = p - q, D_4 = 1 - (p + q), D_5 = -(1 - q) \text{ e } D_6 = 1 - q.$$

O fator Δ usado na integral de superfície é

$$\Delta = (D_1^2 + D_2^2 + D_3^2 + D_4^2 + D_5^2 + D_6^2) = 3 - 4p(1 - p) - 4q(1 - q)$$

A integral de superfície, que será denominada por $H(a_{11}, a_{12}, a_{21}, a_{22})$, vale

$$H(a_{11}, a_{12}, a_{21}, a_{22}) = \int_0^1 \int_0^1 \sqrt{\Delta} p^{a_{1\bullet}-2} (1-p)^{a_{2\bullet}-2} q^{a_{\bullet 1}-2} (1-q)^{a_{\bullet 2}-2} dp dq,$$

em que $a_{i\bullet} = a_{i1} + a_{i2}$ e $a_{\bullet i} = a_{1i} + a_{2i}$.

Utilizando a expansão de Taylor no ponto $(p, q) = (\frac{1}{2}, \frac{1}{2})$ e também algumas propriedades da distribuição Beta, podemos aproximar a função H por

$$H(a_{11}, a_{12}, a_{21}, a_{22}) = 2B(a_{1\bullet}-1; a_{2\bullet}-1)B(a_{\bullet 1}-1; a_{\bullet 2}-1) \left\{ 1 - \frac{a-2}{a-1} \{p^*(1-p^*) + q^*(1-q^*)\} \right\},$$

em que $p^* = \frac{a_{1\bullet}-1}{a-2}$ e $q^* = \frac{a_{\bullet 1}-1}{a-2}$

A função de probabilidade preditiva sob H_0 é dada por:

$$f_0(n_{11}, n_{12}, n_{21}, n_{22}) = \frac{H(A_{11}, A_{12}, A_{21}, A_{22})}{H(a_{11}, a_{12}, a_{21}, a_{22})} \frac{n!}{n_{11}! n_{12}! n_{21}! n_{22}!}$$

Sob H_1 , a função de probabilidade preditiva é

$$f_1(n_{11}, n_{12}, n_{21}, n_{22}) = \frac{B(A_{11}, A_{12}, A_{21}, A_{22})}{B(a_{11}, a_{12}, a_{21}, a_{22})} \frac{n!}{n_{11}! n_{12}! n_{21}! n_{22}!}$$

em que $B(a_{11}, a_{12}, a_{21}, a_{22}) = \frac{\Gamma(a_{11})\Gamma(a_{12})\Gamma(a_{21})\Gamma(a_{22})}{\Gamma(a_{11}+a_{12}+a_{21}+a_{22})}$

De posse dessas funções, pode-se calcular a razão a posteriori:

$$R_{01}(d) = \frac{H(A_{11}, A_{12}, A_{21}, A_{22})}{H(a_{11}, a_{12}, a_{21}, a_{22})} \frac{B(A_{11}, A_{12}, A_{21}, A_{22})}{B(a_{11}, a_{12}, a_{21}, a_{22})}$$

Sendo $H(1, 1, 1, 1) = 2/3$, e considerando prioris uniformes, a expressão da razão de verossimilhanças pode ser simplificada para

$R_{01}(d) = \frac{H(n_{11}+1, n_{12}+1, n_{21}+1, n_{22}+1)}{4} \frac{(n+3)!}{n_{11}!n_{12}!n_{21}!n_{22}!}$, que é aproximadamente, pela expansão de Taylor, igual a

$$R_{01}(d) = \frac{1}{4} \frac{(n+3)(n+2)}{n+1} \frac{\binom{n_{1\bullet}}{n_{11}} \binom{n_{2\bullet}}{n_{21}}}{\binom{n}{n_{1\bullet}}} \left\{ 1 - \frac{n+2}{n+3} [p^{**}(1-p^{**}) + q^{**}(1-q^{**})] \right\}$$

onde $p^{**} = \frac{n_{1\bullet}+1}{n+2}$ e $q^{**} = \frac{n_{\bullet 1}+1}{n+2}$

Note que p^{**} e q^{**} são respectivamente os estimadores de Bayes para p e q .

3.4 Teste de equilíbrio populacional

3.4.1 Definição de equilíbrio populacional

Vamos considerar uma população em que, para determinado par de genes, haja 3 genótipos ligados ao gene X não relacionados ao sexo, isto é, o resultado de um cruzamento possa gerar os genótipos XX, Xx e xx. Sejam p_1, p_2 e p_3 respectivamente as proporções desses genótipos na população.

O interesse dos geneticistas é verificar se a população em estudo está em equilíbrio, isto é, se segue a lei de Hardy-Weinberg que diz que em uma população em equilíbrio as proporções genotípicas satisfazem a seguinte propriedade:

$\exists p \in (0, 1)$ tal que $p_1 = p^2$, $p_2 = 2p(1 - p)$ e $p_3 = (1 - p)^2$

3.4.2 O teste clássico

Vários testes podem ser empregados para testar o equilíbrio de Hardy-Weinberg, entre eles o χ^2 , o teste exato condicional e o teste da razão de verossimilhanças.

Maistre (1993) comparou esses testes e concluiu que o teste exato condicional tem uma melhor desempenho, principalmente no caso de amostras pequenas. Portanto, esse teste será utilizado neste caso.

As amostras são ordenadas de acordo com as suas probabilidades multinomais, sob a hipótese de equilíbrio e condicionadas ao número total de genes de determinado tipo observados.

Portanto, vamos considerar a amostra $d = (n_1, n_2, n_3)$, onde n_1 é o número de pessoas na amostra com genótipo XX, n_2 é o número de pessoas na amostra com genótipo Xx e n_3 é o número de pessoas na amostra com genótipo xx.

Seja n_i o número total de genes do tipo i na amostra. Então, $n_1 = 2n_1 + n_2$, isto é, n_1 é o número total de genes X na amostra.

A probabilidade de a amostra (n_1, n_2, n_3) ser selecionada sob um processo de seleção aleatório condicionando-se no total de gene X é

$$Pr\{(n_1, n_2, n_3)|n_1\} = \frac{n!2^{n_2}(2n_1+n_2)!(n_2+2n_3)!}{2n!n_1!n_2!n_3!}$$

O p-value é calculado somando-se as probabilidades das amostras com o mesmo número de genes e com probabilidade menor do que a amostra observada. O teste exato condicional é análogo ao Teste Exato de Fisher para tabelas de contingência.

3.4.3 O teste de Bayes

A partição do espaço paramétrico que define a hipótese de equilíbrio em uma população pode ser expressa da seguinte forma:

$$\Omega_0 = \{(x, y, z); x \in (0, 1), y = 2\sqrt{x}(1 - \sqrt{x}), z = (1 - \sqrt{x})^2\}$$

$$\Omega_1 = \{(x, y, z); x \geq 0; y \geq 0; z \geq 0 \text{ e } x + y + z = 1\}$$

A verossimilhança vale:

$$L(\omega|d) = p_1^{n_1} p_2^{n_2} p_3^{n_3} h(n_1, n_2, n_3),$$

em que $\omega = (p_1, p_2, p_3)$.

Se considerarmos uma distribuição multinomial, $h(n_1, n_2, n_3) = \frac{1}{n_1! n_2! n_3!}$

A densidade a priori escolhida é a Dirichlet, por ser da família de distribuições conjugadas da exponencial e será denotada por $Dir_3(a_1, a_2, a_3)$, isto é,

$$g(\omega) = \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} p_1^{a_1-1} p_2^{a_2-1} p_3^{a_3-1}.$$

Utilizando as propriedades da distribuição Dirichlet, a distribuição a posteriori é:

$$g(\omega|d) \sim Dir_3(a_1 + n_1, a_2 + n_2, a_3 + n_3) = Dir_3(A_1, A_2, A_3)$$

Similarmente ao teste de independência, é necessário achar as derivadas do sistema de equações em Ω_0 .

O vetor de derivadas é $(D_1, D_2, D_3) = (2p, 2 - 4p, 2p - 2)$

O fator de cálculo para a integral de linha em Ω_0 será

$$\int_0^1 \sqrt{D_1^2 + D_2^2 + D_3^2} dp = 2\sqrt{2} \int_0^1 \sqrt{1 - 3p(1 - p)} dp = 1.9519$$

Para simplificar, a seguinte notação será utilizada nos cálculos abaixo:

$$B(a_1, a_2, a_3) = \frac{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)}{\Gamma(a_1+a_2+a_3)} \text{ e}$$

$$E(a_1, a_2, a_3) = \int_0^1 \sqrt{1 - 3p(1-p)} p^{2a_1+a_2-3} (1-p)^{2a_3+a_2-3} dp$$

A função preditiva dos dados sob H_0 é

$$f_0(d) = \frac{\int L(\omega|d)g(\omega)d\Omega_0}{\int g(\omega)d\Omega_0}, \text{ em que:}$$

$$\int g(\omega)d\Omega_0 = 2^{a_2} \sqrt{2} \frac{E(a_1, a_2, a_3)}{B(a_1, a_2, a_3)},$$

$$\int g(\omega)L(\omega|d)d\Omega_0 = 2^{A_2} \sqrt{2} \frac{E(A_1, A_2, A_3)}{B(a_1, a_2, a_3)} h(n_1, n_2, n_3)$$

Deve-se observar que

$$\int g(\omega)d\Omega_1 = 1 \text{ e}$$

$$\int g(\omega)L(\omega|d)d\Omega_1 = \frac{B(A_1, A_2, A_3)}{B(a_1, a_2, a_3)} h(n_1, n_2, n_3)$$

Portanto,

$$f_0(d) = 2^{n_2} \frac{E(A_1, A_2, A_3)}{E(a_1, a_2, a_3)} h(n_1, n_2, n_3) \text{ e}$$

$$f_1(d) = \frac{B(A_1, A_2, A_3)}{B(a_1, a_2, a_3)} h(n_1, n_2, n_3)$$

A razão de verossimilhanças pode então ser escrita como:

$$R_{01}(d) = 2^{n_2} \frac{E(A_1, A_2, A_3)}{E(a_1, a_2, a_3)} \frac{B(A_1, A_2, A_3)}{B(a_1, a_2, a_3)}$$

No caso de uma priori $Dir_3(1, 1, 1)$, tem-se:

$$f_0(d) = n!2^{n_2} \frac{\int_0^1 \sqrt{1-3p(1-p)} p^{2n_1+n_2} (1-p)^{2n_3+n_2} dp}{\int_0^1 \sqrt{1-3p(1-p)} dp} h(n_1, n_2, n_3) \text{ e}$$

$$f_1(d) = \frac{2}{(n+1)!(n+2)!} n_1! n_2! n_3! h(n_1, n_2, n_3)$$

Finalmente, temos que

$$R_{01}(d) = \frac{E(n_1+1, n_2+1, n_3+1)}{n_1! n_2! n_3!} \frac{(n+2)! 2^{n_2-1}}{0.69}$$

Para calcular a razão de verossimilhanças, é necessário fazer uma aproximação pelo polinômio de Taylor. Tem-se então a seguinte expressão:

$$R_{01}(d) = \frac{(n+2)! 2^{n_2-1}}{n_1! n_2! n_3!} \frac{(2n_1+n_2)!(2n_3+n_2)!}{3*(2n+1)!} \left[5 - 12 \frac{(2n_1+n_2+1)(2n_3+n_2+1)}{(2n+2)(2n+3)} \right]$$

Capítulo 4

Comparação entre as medidas de evidência

4.1 Exemplos de inconsistência no valor do p-value clássico

Serão apresentados agora alguns exemplos cujo cálculo do p-value clássico leva a conclusões equivocadas quando a hipótese alternativa não é considerada.

Exemplo 1: Seja X um variável aleatória com distribuição normal com média zero e variância desconhecida.

Vamos supor que queiramos testar:

$$H_0: \sigma = 2 \text{ vs. } H_A: \sigma = 1$$

O p-value usual é calculado da seguinte maneira:

$$p(t) = P(T \geq t | \sigma = 2) = 2\Phi(-\frac{\sqrt{t}}{2}),$$

em que $T=T(X)=X^2$ é uma estatística mínima suficiente e Φ é a função de distribuição normal padronizada

Portanto, se $t = x^2 = 16$, rejeitamos H_0 com nível descritivo $p(16)=0.0454$, concluindo em favor de H_A .

O problema aqui é que a probabilidade de observar $t=16$ sob H_A é muito menor, já que o p-value é calculado como

$$p_A(t) = P(T \geq t | \sigma = 1) = 2\Phi(-\sqrt{t}).$$

Sendo assim, o valor de $p_A(t)$ é 0.00003 para $t = 16$.

Neste caso, o p-value leva a rejeitar H_0 , mesmo quando a evidência contra a hipótese alternativa é mais forte.

Consideremos agora o P-value, que no caso de hipóteses simples é definido como

$$P(r) = P(R \geq r | H),$$

em que $R = \frac{f_A(X)}{f_H(X)}$ é a razão de verossimilhanças.

Por esse método, o P-value no ponto $t = 16$ é

$$P(2e^{-6}) = 1 - 2\Phi(-2) = 0.9546,$$

o que nos leva a não rejeitar H_0 .

Exemplo 2: Uma observação x de uma variável aleatória que tem função densidade f é obtida. Queremos testar:

$$H_0: f \text{ é } N(0;4)$$

$H_A : f \text{ é Cauchy-padrão}$

Se observarmos $t = x^2 = 16$ como no exemplo anterior, teremos o mesmo p-value, já que a hipótese alternativa não é considerada.

A diferença aqui é que $p_A(16) = 0.156$, valor maior que o p-value sob H_0 .

O valor resultante do P-value, fazendo-se a ordenação das razões de verossimilhanças, é $P(16) = 0.55$, favorecendo H_0 .

Nos dois exemplos anteriores, a hipótese alternativa era simples. Consideremos agora o teste

$H_0: \sigma = 2 \text{ vs. } H_A : \sigma \neq 2$

Como agora temos uma hipótese composta, precisamos definir uma distribuição a priori no conjunto dos pontos do espaço paramétrico que pertencem à hipótese alternativa. Vamos supor que a priori para $1/\sigma^2$ é uma χ^2 com 1 grau de liberdade. Temos então que f_A é uma Cauchy-padrão e o cálculo do p-value é feito como no exemplo 2.

4.2 Comparação entre as evidências obtidas pelos dois métodos

4.2.1 Teste de binomiais

Para o cálculo do P-value foram consideradas prioris uniformes para as proporções.

Nos gráficos 1.1, 1.2, 1.3, 2.1, 2.2 e 2.3 (ver apêndice), foi calculado o p-value pelos três métodos para cada elemento do espaço amostral e para tamanhos de amostra de cada binomial igual a 10 e 60.

Nota-se que os três gráficos são semelhantes. Um aspecto interessante é que, enquanto o teste Exato de Fisher e o qui-quadrado têm p-value próximo de 1 em toda a diagonal, o P-value tem uma pequena "queda" quando os valores observados estão próximos de 0.5.

Uma pergunta que se pode fazer é: porque o P-value é menor na área central do gráfico, mesmo quando as proporções observadas são iguais?

A resposta está na razão de verossimilhanças, que é proporcional a

$$R_{01}(x, y) = \frac{\binom{m}{x} \binom{n}{y}}{\binom{m+n}{x+y}}$$

Quando as proporções observadas são iguais ($x = y$ no caso de $m = n$), é fácil ver que a razão de verossimilhança é maior nas extremidades; com isso, o P-value é maior. Esse comportamento pode ser visto no gráfico 1.1.

Uma diferença importante entre o teste exato de Fisher e o P-value é que o último é a soma de f_0 de todas as tabelas com apenas uma marginal fixa, enquanto que o teste exato de Fisher as probabilidades somadas são condicionadas a tabelas com todas as marginais fixas.

Já a diferença nos valores encontrados no P-value e no teste qui-quadrado acontece nos pontos onde a proporção observada é próxima de 0 ou 1. Nesses casos, a aproximação pela distribuição qui-quadrado é ruim, já que a frequência observada é pequena.

No geral, as diferenças encontradas nos valores do P-value e dos outros dois testes são grandes apenas quando a hipótese nula não é rejeitada. Isto pode ser visto nos gráficos 1.5, 1.6, 2.5 e 2.6. A maioria dos pontos está na diagonal (quando os valores do p-value e do P-value Bayesiano coincidem). Nas tabelas abaixo, pode ser visto que os diferentes métodos rejeitam a hipótese nula (com nível de significância de 5%) para quase todos os mesmos pontos amostrais. No corpo das tabelas, é mostrado o número de pontos amostrais onde rejeita-se ou não a hipótese, para ambos os testes.

		χ^2		T.E.Fisher	
		Não rejeita H_0	Rejeita H_0	Não rejeita H_0	Rejeita H_0
P-value	Não rejeita H_0	71	0	71	0
	Rejeita H_0	4	46	16	34

Tamanho da amostra: $m = n = 10$

		χ^2		T.E.Fisher	
		Não rejeita H_0	Rejeita H_0	Não rejeita H_0	Rejeita H_0
P-value	Não rejeita H_0	975	20	995	0
	Rejeita H_0	36	2690	132	2594

Tamanho da amostra: $m = n = 60$

Nos três casos, com o aumento do tamanho da amostra, os padrões dos gráficos se mantêm, exceto pelo fato que a região de rejeição da hipótese nula fica mais abrangente.

4.2.2 Comparação de duas Poisson

O P-value aqui não é calculado com precisão. Para tanto, precisaríamos calcular a função densidade em todos os pontos amostrais, o que é impossível, já que há infinitos pontos. Portanto, o P-value foi calculado dentro de um conjunto limitado de pontos, que englobam cerca de 99% da massa de probabilidade conjunta. Sendo assim todos os P-values Bayesianos apresentados nos gráficos 3.1 e 3.3 (ver apêndice) têm um erro de 1%.

No capítulo anterior, foi considerada a distribuição Gama para λ e μ , pois esta última faz parte da classe de distribuições conjugadas para Poisson. A distribuição a priori considerada será então a Gama (1,1), que equivale a uma exponencial com média 1. Note que os níveis descritivos encontrados pelos dois métodos apresentam comportamento diferente (ver gráficos 3.1 a 3.4 no apêndice). Para comparar os dois métodos, seria interessante considerar uma priori uniforme para o parâmetro da Poisson. Um ponto interessante é que mesmo

considerando priori uniforme o comportamento do P-value Bayesiano não muda muito. Isso se deve ao fato de a priori ter pouca influência na distribuição a posteriori quando o tamanho da amostra é grande.

4.2.3 Teste de independência

Para o cálculo do P-value Bayesiano, foi considerada uma priori Dirichlet(1,1,1). Os gráficos apresentados no apêndice B mostram um comportamento semelhante entre o p-value qui-quadrado e o P-value. Deve-se notar entretanto que o p-value calculado pela distribuição qui-quadrado é mais conservador, pois geralmente apresenta valor mais alto que o P-value Bayesiano.

Isto também pode ser visto pela tabela a seguir:

		χ^2	
		Não rejeita H_0	Rejeita H_0
P-value Bayesiano	Não rejeita H_0	512	16
	Rejeita H_0	68	220

Tamanho da amostra: $n = 15$

		χ^2	
		Não rejeita H_0	Rejeita H_0
P-value Bayesiano	Não rejeita H_0	15357	428
	Rejeita H_0	1602	22324

Tamanho da amostra: $n = 60$

Pode-se notar que teste chi-quadrado rejeita menos vezes H_0 do que o P-value Bayesiano para os dois tamanhos de amostra.

4.2.4 O teste de equilíbrio de Hardy-Weinberg

A diferença entre os dois métodos tem comportamento semelhante ao caso do teste de binomiais (ver gráficos no apêndice B); quando $\hat{p}$ e $\hat{q}$ são próximos de 0.5, o P-value, apesar de não rejeitar H_0 , não apresenta um P-value tão alto quanto o apresentado pelo teste exato condicional. Isto acontece novamente devido ao fato do P-value considerar o espaço amostral inteiro e ordenar as razões de verossimilhanças, que são maiores nos extremos. Já o teste exato condicional a probabilidade de observar a amostra no valor da estatística $T = 2n_1 + n_2$, o que restringe o espaço amostral.

Apesar dessa diferença, observa-se na tabela a seguir que os métodos rejeitam ou não H_0 (com nível de significância 5%) para os mesmos pontos amostrais, em sua maioria.

		Teste exato	
		Não rejeita H_0	Rejeita H_0
P-value	Não rejeita H_0	109	0
	Rejeita H_0	28	94

Tamanho da amostra: $n = 20$

		Teste exato	
		Não rejeita H_0	Rejeita H_0
P-value	Não rejeita H_0	552	2
	Rejeita H_0	106	1231

Tamanho da amostra: $n = 60$

4.3 Comparação entre os erros reportados pelos dois métodos

As probabilidades dos erros do tipo I (rejeitar H_0 quando esta é verdadeira) e do tipo II (não rejeitar H_0 quando esta é falsa) são calculados com base em suposições e afirmações teóricas usadas na construção do teste. Entretanto, os valores dos erros que efetivamente são cometidos podem ser diferentes dos teóricos. Sendo assim, será feita uma simulação para estimar esses erros. A comparação entre os métodos será baseada na seguinte definição:

Definição: Um teste δ_1 é superior a um teste δ_2 se as duas condições abaixo forem satisfeitas:

i) $\alpha^*(\delta_1) + k\beta^*(\delta_1) \leq \alpha^*(\delta_2) + k\beta^*(\delta_2)$

ii) $|\alpha(\delta_1) - \alpha^*(\delta_1)| \leq |\alpha(\delta_2) - \alpha^*(\delta_2)|$ e $|\beta(\delta_1) - \beta^*(\delta_1)| \leq |\beta(\delta_2) - \beta^*(\delta_2)|$, em que:

k depende da relação de importância entre os dois erros (nesse caso, vale 1);

α^* é a estimativa da proporção esperada de vezes que rejeitaríamos H_0 quando esta é verdadeira;

β^* é a estimativa da proporção esperada de vezes que não rejeitaríamos H_0 quando esta é falsa;

α e β são as probabilidades dos erros provenientes dos testes.

As duas condições da definição anterior indicam que o melhor teste é aquele com menor erro total e cujos valores dos erros efetivos estão mais próximos dos valores calculados pelo estatístico.

Deve-se ressaltar que a probabilidade do erro β dos testes não-Bayesianos são desconhecidos.

A seguir, são apresentados os resultados de simulações feitas para os testes de hipóteses apresentados no capítulo 2.

O cálculo de α e β é feito da seguinte maneira:

$$\alpha = \sum_C f_0 \text{ e}$$

$$\beta = 1 - \sum_C f_1,$$

onde C é a região crítica do teste.

4.3.1 Teste de binomiais

Para calcular α^* , foram considerados os seguintes pontos em Ω_0 : $p = q = \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$, onde n é o tamanho da amostra, e foram geradas 100 amostras de tamanho 10 e 60 para cada um desses pontos. Foram contados os pontos com evidência menor que 0.05.

Para calcular β^* , foram considerados os seguintes pontos em Ω_1 : $p, q = \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$, $p \neq q$, onde n é o tamanho da amostra. Foram contadas as amostras com evidência menor que 0.05. No corpo das tabelas a seguir são apresentadas as probabilidades de cada erro, para os testes estudados.

	Simulado		Teórico	
	α	β	α	β
P-value	0.044	0.490	0.050	0.587
χ^2	0.021	0.524	0.050	-
T.E.Fisher	0.006	0.600	0.050	-

Tamanho da amostra: $n = 10$

	Simulado		Teórico	
	α	β	α	β
P-value	0.052	0.251	0.050	0.267
χ^2	0.049	0.254	0.050	-
T.E.Fisher	0.027	0.286	0.050	-

Tamanho da amostra: $n = 60$

As duas tabelas anteriores mostram que o teste de Bayes apresenta probabilidade dos erros mais próximos da probabilidade teórica, mesmo para amostras pequenas. No caso do qui-quadrado, a probabilidade de erro só se aproxima da probabilidade teórica quando a amostra é grande. O teste exato de Fisher não se aproxima nem com amostra de tamanho 60. Isto se deve ao fato das marginais serem fixas; o p-value encontrado é condicional a essas marginais. Sendo assim o espaço considerado por este teste não é o espaço amostral resultante do experimento. A soma das probabilidades dos erros também é menor no teste de Bayes.

4.3.2 Teste de Poisson

Para calcular a probabilidade de erro α , foram geradas observações de uma distribuição exponencial de parâmetro 1 para obter valores do parâmetro λ . Dado o valor deste parâmetro, são geradas duas observações x e y com distribuição Poisson(λ) e então é calculado o P-value Bayesiano e o p-value utilizando o teste condicional. Foram geradas 100 amostras para x e y .

A probabilidade de erro β é obtida de modo semelhante ao descrito acima, mas agora um outro parâmetro μ é obtido de uma distribuição exponencial de parâmetro 1. Agora, x tem distribuição Poisson(λ) e y tem distribuição Poisson(μ). No corpo das tabelas a seguir, é mostrada a probabilidade de cada erro para os testes estudados.

	Simulado		Teórico	
	α	β	α	β
P-value	0.084	0.457	0.050	0.542
Condicional	0.036	0.512	0.050	-

Tamanho da amostra: $n = 10$

	Simulado		Teórico	
	α	β	α	β
P-value	0.071	0.285	0.050	0.288
Condicional	0.041	0.318	0.050	-

Tamanho da amostra: $n = 60$

4.3.3 Teste de independência

Foram simuladas 100 amostras de tamanho 10 e 60 para cada tabela gerada pelo seguinte algoritmo:

Ω_0 :

$p = 0.5, 0.51, \dots, 0.99$

$p_{11} = y * p, y = 0.5, \dots, 1$

$p_{12} = p - p_{11};$

$p_{21} = y * (1 - p)$

$p_{22} = (1 - p) - p_{21}$

Ω_1 :

$$p_{11} = 0, 0.1, 0.2, \dots, 0.9, 1$$

$$p_{12} = 0, 0.1, \dots, 1 - p_{11};$$

$$p_{21} = 0, 0.1, \dots, 1 - p_{11} - p_{12}$$

$$p_{22} = 1 - p_{11} - p_{12} - p_{21}$$

$$\text{e } p_{11} * p_{22} \neq p_{12} * p_{21}$$

Foram contadas as amostras com evidência menor que 0.05 geradas em Ω_0 e Ω_1 . O resultado é apresentado nas tabelas a seguir:

	Simulado		Teórico	
	α	β	α	β
P-value	0.056	0.598	0.048	0.647
χ^2	0.036	0.646	0.050	-

Tamanho da amostra: $n = 15$

	Simulado		Teórico	
	α	β	α	β
P-value	0.055	0.382	0.050	0.397
χ^2	0.043	0.407	0.050	-

Tamanho da amostra: $n = 60$

As tabelas anteriores mostram que o P-value Bayesiano é superior, pois a probabilidade do erro reportado é mais próxima do teórico e a soma das probabilidades dos erros é menor do que a soma das probabilidades dos erros do qui-quadrado.

4.3.4 Teste de Hardy-Weinberg

Foram simuladas 100 amostras de tamanho 20 e 60 para cada tabela gerada pelo seguinte algoritmo:

Ω_0 :

$$x_1 = 0, 0.01, \dots, 0.99, 1$$

$$x_3 = (1 - \sqrt{x_1})^2;$$

$$x_2 = 1 - x_1 - x_3$$

Ω_1 :

$$x_1 = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1$$

$$x_2 = 0, \frac{1}{n}, \frac{2}{n}, \dots, n - x_1;$$

$$\text{e } x_3 \neq (1 - \sqrt{x_1})^2$$

Foram contadas as amostras com evidência menor que 0.05 geradas em Ω_0 e Ω_1 . No corpo das tabelas a seguir é mostrada a probabilidade dos erros:

	Simulado		Teórico	
	α	β	α	β
P-value	0.063	0.458	0.047	0.472
Teste exato	0.015	0.572	0.050	-

Tamanho da amostra: $n = 20$

	Simulado		Teórico	
	α	β	α	β
P-value	0.045	0.292	0.050	0.293
Teste exato	0.020	0.347	0.050	-

Tamanho da amostra: $n = 60$

As tabelas anteriores mostram que o P-value Bayesiano apresenta superioridade sobre o teste exato condicional. Este último tem o mesmo problema do teste exato de Fisher: o p-value é condicionado aqui à estatística $2n_1 + n_2$. Por esse motivo, a probabilidade encontrada através de simulação ainda está longe da teórica mesmo com tamanho da amostra igual a 60. Também nota-se que a soma das probabilidades dos erros gerados pelo P-value Bayesiano é menor.

Conclusão

O P-value tem várias vantagens sobre os métodos clássicos apresentados: leva a hipótese alternativa em consideração, evitando inconsistências; pode ser aplicado a amostras pequenas por não necessitar de aproximações assintóticas; minimiza a soma dos erros cometidos e possui probabilidade de erro próxima à probabilidade teórica.

A única desvantagem é que, devido à falta de fórmula fechada, o P-value necessita de um tempo computacional grande, proporcional ao tamanho da amostra. Quando a amostra é grande, pode-se usar o teste qui-quadrado, já que os valores de ambos os testes rejeitam a hipótese nula nos mesmos casos. Sendo assim, é uma alternativa aos métodos clássicos disponíveis.

Apêndice A

Aspectos Computacionais

A.1 Simulação

Os programas para calcular os valores dos p-values foram feitos no software MATLAB. Há dois tipos distintos de programas. A lógica de cada um depende do objetivo que se tem. Se o objetivo for calcular o P-value para uma determinada amostra, o programa é feito de uma maneira. Se o objetivo for calcular o P-value para todas as amostras possíveis provenientes de um espaço amostral, o programa deve ser feito de outra maneira para que haja maior eficiência computacional.

Os passos para se calcular um P-value são os seguintes:

- 1) calcular a razão de verossimilhança e f_0 para todos os valores amostrais possíveis, inclusive para a amostra observada;
- 2) somar f_0 para todos os pontos amostrais cuja razão de verossimilhança é menor do que a razão verossimilhança observada.

O exemplo abaixo mostra como seria o cálculo no caso do P-value binomial para duas

amostras de tamanho m e n e x e y sucessos, respectivamente:

```
rvobs=vero_bin(m,n,x,y);
soma=0;
for d1=0:m
    for d2=0:n
        rv=vero_bin(m,n,d1,d2);
        if rv<=rvobs
            f0=nchoosek(m,d1)*nchoosek(n,d2)*beta(d1+d2+1,m+n-d1-d2+1)/beta(1,1);
            soma=soma+f0;
        end
    end
end
result=soma;
```

Note que para o cálculo de um P-value, todas as razões de verossimilhança são calculadas. No caso das simulações, em que todos os P-values são calculados, a rotina acima seria calculada várias vezes. Isso tornaria o programa ineficiente, pois todas as razões de verossimilhança seriam calculadas cada vez que a rotina fosse chamada.

Por outro lado, uma rotina que armazenasse todas as razões de verossimilhança e todos os f_0 em uma matriz tornaria o cálculo mais rápido. A lógica é descrita no programa a seguir, onde as razões de verossimilhança são guardadas na coluna 1 da matriz $n \times 2$ e f_0 é guardado na coluna 2:

```
for i=1:n
    rvobs=matriz(i,1);
    pval=0;
    for j=1:n
        if matriz(j,1)<=rvobs
            pval=pval+vet_matriz(j,2);
        end
    end
end
```

```

end
vet(i)=pval;
end
end

```

Os P-values são, então, guardados em um vetor.

Um ponto interessante é que em alguns casos a razão de verossimilhança não precisa ser calculada para todo o espaço amostral, pois elementos diferentes do espaço amostral produzem a mesma verossimilhança, como no exemplo abaixo, no teste de independência:

2	3	5	3	2	5	1	6	7	6	1	7
1	6	7	6	1	7	2	3	5	3	2	5
3	9	12	9	3	12	3	9	12	9	3	12

Sendo assim, no exemplo anterior o p-value pode ser calculado considerando-se em torno de 1/4 do espaço amostral e o seu valor pode ser multiplicado por 4 na maioria dos casos, o que reduz em aproximadamente 75% o tempo computacional usual.

A.2 Cálculo da verossimilhança

A limitação numérica dos computadores pessoais não permite que sejam feitas contas com valores que ultrapassem $170! \approx 10^{300}$. Sendo assim, o tamanho da amostra fica muito limitado.

Uma maneira de contornar esse problema é utilizar o logaritmo. No exemplo abaixo, que mostra a razão de verossimilhança da tabela de independência, pode-se ver a simplificação:

$$\ln(R_{01}(d)) = cte + \ln\binom{n_{1\bullet}}{n_{11}} + \ln\binom{n_{2\bullet}}{n_{22}} - \ln\binom{n}{n_{1\bullet}} + \ln\left(1 - \frac{(n+2)}{(n+3)}[p^{**}(1-p^{**}) + q^{**}(1-q^{**})]\right)$$

onde:

$$cte = \ln\left(\frac{1}{4}\right) + \ln\left(\frac{(n+3)(n+2)}{(n+1)}\right)$$

$$p^{**} = \frac{n_{1\bullet}+1}{n+2} \text{ e } q^{**} = \frac{n_{\bullet 1}+1}{n+2}$$

As combinações pode ser simplificadas ainda mais:

$$\ln\binom{n_{1\bullet}}{n_{11}} + \ln\binom{n_{2\bullet}}{n_{22}} - \ln\binom{n}{n_{1\bullet}} = \sum_{i=n_{11}+1}^{n_{1\bullet}} \ln(i) + \sum_{i=n_{12}+1}^{n_{\bullet 2}} \ln(i) + \sum_{i=n_{21}+1}^{n_{\bullet 1}} \ln(i) + \sum_{i=n_{22}+1}^{n_{2\bullet}} \ln(i) + cte$$

Dessa maneira, as contas se reduzem a logaritmos e somas. A transformação de volta nem precisa ser feita, pois o logaritmo sendo uma função crescente mantém a ordenação da razão de verossimilhança.

A.3 Tempo de computação

O caso que será analisado é o caso do teste de independência. Dentre os 4 testes, este é o que mais demanda tempo computacional, pois o número de pontos amostrais em uma tabela 2x2 é proporcional ao cubo do tamanho da amostra. Em um Pentium 300 Mhz foi observado o seguinte tempo de computação:

Tamanho da amostra	10	50	100	200
Tempo de computação (min)	0.02	1.2	8.8	67.6

No geral, o tempo computacional para tabelas de independência $s \times r$ é proporcional a n^{s*r-1} .

Apêndice B

Gráficos

TESTE DE DUAS BINOMIAIS

Nos gráficos abaixo: Eixos x e y: proporções observadas. Eixo z (cores): valor da evidência

Tamanho da amostra: $m=n=60$

Gráfico 1.1: P-value Bayesiano

Gráfico 1.2: p-values calculados pelo teste chi-quadrado

Gráfico 1.3: p-value do teste exato de Fisher

Gráfico 1.4: Escala de cores do valor da evidência

Gráfico 1.5: gráfico de dispersão: P-value Bayesiano vs. p-value chi-quadrado

Gráfico 1.6: gráfico de dispersão: P-value Bayesiano vs. p-value TE Fisher

TESTE DE BINOMIAIS

Nos gráficos abaixo: Eixos x e y: proporções observadas. Eixo z (cores): valor da evidência

Tamanho da amostra: $m=n=10$

Gráfico 2.1: P-value Bayesiano

Gráfico 2.2: p-values calculados pelo teste chi-quadrado

Gráfico 2.3: p-value do teste exato de Fisher

Gráfico 2.4: Escala de cores do valor da evidência

Gráfico 2.5: gráfico de dispersão: P-value Bayesiano vs. p-value chi-quadrado

Gráfico 2.6: gráfico de dispersão: P-value Bayesiano vs. p-value TE Fisher

TESTE DE POISSON

Tamanho da amostra: $n=m=10$

Gráfico 3.1 - Valores do P-value Bayesiano

Tamanho da amostra: $n=m=60$

Gráfico 3.3 - Valores do P-value Bayesiano

Gráfico 3.2 - Valores do teste condicional

Gráfico 3.4 - Valores do teste condicional

TESTE DE INDEPENDÊNCIA

Tamanho da amostra: n=15

x11	x12	
0	x22	15

x11	x12	
3	x22	15

x11	x12	
7	x22	15

Gráfico 4.1

Gráfico 4.3

Gráfico 4.5

Gráfico 4.2

Gráfico 4.4

Gráfico 4.6

Gráfico 4.7 - Escala de cores da evidência, valor representado em porcentagem

TESTE DE INDEPENDÊNCIA

Tamanho da amostra: n=60

x11	x12
0	x22
	60

x11	x12
15	x22
	60

x11	x12
30	x22
	60

Gráfico 5.1

Gráfico 5.3

Gráfico 5.5

Gráfico 5.2

Gráfico 5.4

Gráfico 5.6

Gráfico 5.7 - Escala de cores da evidência, valor representado em porcentagem

TESTE DO EQUILÍBRIO DE HARDY-WEINBERG

Tamanho da amostra: $n=20$

Gráfico 6.1 - P-value bayesiano para o teste de Hardy-Weinberg

Gráfico 6.2 - p-value calculado pelo teste exato condicional

Gráfico 6.5 - Escala de cores da evidência

Tamanho da amostra: $n=60$

Gráfico 6.3 - P-value bayesiano para o teste de Hardy-Weinberg

Gráfico 6.4 - p-value calculado pelo teste exato condicional

Referências Bibliográficas

- Basu, D. (1977). On the elimination of nuisance parameters, *Journal of The American Statistical Association* **72**: 355–366.
- Courant, R. & John, F. (1974). *Introduction to Calculus and Analysis*, Vol. 2, John Wiley, New York.
- Cox, D. R. (1958). Some problems connected with statistical inference, *Annals of Mathematical Statistics* **29**: 357–372.
- DeGroot, M. H. (1989). *Probability and Statistics*, 2 edn, Addison Wesley, Menlo Park.
- Irony, T. Z. & Pereira, C. A. B. (1995). Using surface integrals to distribute prior information among the hypotheses, *Resenhas IME-USP* **2**: 27–46.
- Maistre, P. J. (1993). *Comparison of Statistical Tests for Independence at Genetic Loci with Many Alleles*, PhD thesis, NCSU, Raileigh, NC.
- Pereira, C. A. B. (1985). *Teste de Hipóteses Definidas em Espaços de Diferentes Dimensões: Visão Bayesiana e Interpretação Clássica*, Tese de Doutorado, IME-USP.
- Pereira, C. A. B. & Wechsler, S. (1993). On the concept of p-value, *Brazilian Journal of Probability and Statistics* **7**: 159–177.