

Teste de Significância Bayesiano no Problema de Behrens-Fisher

Pedro Silvestre da Silva Campos,

Maria Regina Madruga,

Héilton Ribeiro Tavares

Programa de Pós-Graduação em Matemática e Estatística, CCEN, UFPA,

Campus do Guamá, 66075-110, Belém, PA

E-mail: psscama@yahoo.com.br, madruga@ufpa.br,

heliton@ufpa.br

1 Introdução

Consideremos $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ variáveis aleatórias com médias μ_x e μ_y desconhecidas e variâncias σ_x^2 e σ_y^2 desconhecidas.

O problema de *Behrens-Fisher* consiste em testar $H_0 : \mu_x = \mu_y$ contra as alternativas $\mu_x < \mu_y$, $\mu_x > \mu_y$ ou $\mu_x \neq \mu_y$. Para este trabalho consideraremos o teste para as hipóteses

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x \neq \mu_y,$$

supondo que as variâncias populacionais σ_x^2 e σ_y^2 são diferentes e desconhecidas. Neste caso, o parâmetro de interesse é $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$, com o espaço paramétrico dado por $\Theta = \{\theta : -\infty < \mu_x, \mu_y < +\infty, \sigma_x, \sigma_y > 0\}$, e a hipótese H_0 determinando o subconjunto $\Theta_0 = \{\theta \in \Theta : \mu_x = \mu_y\}$.

Na abordagem estatística Clássica este problema tem algumas soluções aproximadas, que são discutidas e comparadas em Mehta e Srinivasan(1970). Agora, se as duas variâncias forem iguais ou se a razão entre elas for conhecida, existem soluções exatas do ponto de vista Clássico (ver Lehmann(1959)). Os procedimentos Bayesianos usuais para testes de hipóteses não se aplicam no problema de Behrens-Fisher, talvez porque as prioris usuais quando se tem verossimilhanças normais sejam impróprias.

Devido à complexidade para a solução deste problema, tem surgido algumas propostas para solução do mesmo. Neste trabalho será usado o teste de significância Bayesiano proposto por Pereira e Stern (1999), cujo procedimento é denominado **Full Bayesian Significance Test (FBST)**, cuja definição é apresentada a seguir

Definição 1. Considere um modelo estatístico paramétrico, i.e., uma quintupla $(\mathcal{X}, \mathcal{A}, \mathbf{F}, \Theta, \pi)$, onde \mathcal{X} é um espaço amostral, \mathcal{A} é uma σ -álgebra conveniente de subconjuntos de \mathcal{X} , \mathbf{F} é uma classe de distribuições de probabilidade em \mathcal{A} indexadas no espaço paramétrico Θ e π é uma densidade a priori em (uma σ -álgebra de) Θ . Suponha que um subconjunto Θ_0 de Θ tendo medida

de Lebesgue nula (com respeito a Θ) é de interesse. Seja $\pi(\theta|\mathbf{x})$ uma densidade posterior de θ , dada a observação amostral \mathbf{x} , e $T(\mathbf{x}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) > \sup_{\Theta_0} \pi(\theta|\mathbf{x})\}$. A **medida de evidência de Pereira-Stern** é definida como $EV(\Theta_0, \mathbf{x}) = 1 - Pr[\theta \in T(\mathbf{x})|\mathbf{x}]$ e um teste (ou procedimento) de Pereira-Stern é aceitar Θ_0 sempre que $EV(\Theta_0, \mathbf{x})$ é "grande".

Eles propuseram uma medida de evidência em favor de uma hipótese precisa (hipótese com medida de Lebesgue nula), que considera todos os pontos do espaço paramétrico cujos valores da densidade posterior são, no máximo, tão grandes quanto seu supremo em Θ_0 ; falando grosseiramente, considera todos os pontos que são menos "prováveis" do que algum ponto em Θ_0 . De acordo com Pereira e Stern (1999), um valor grande da $EV(\Theta_0, \mathbf{x})$ significa que o subconjunto Θ_0 cai em uma região do espaço paramétrico de alta probabilidade posterior, portanto, os dados suportam a hipótese nula H_0 . Madruga et al. (2001) apresentam funções de perda que tornam o procedimento de Pereira-Stern um legítimo procedimento "Bayesiano", pois devem ser chamados "Bayesianos" apenas os procedimentos que minimizam funções perdas esperadas, a solução coerente para o problema de decisão. Propriedades de invariância da $EV(\Theta_0, \mathbf{x})$ são discutidas em Madruga et al. (2003).

Nosso objetivo é comparar esta medida de evidência do FBST com a medida de evidência Clássica, o p -value do **Teste da Razão de Verossimilhança Generalizada (TRVG)**. A definição do TRVG (Bolfarine e Sandoval (2001)) é apresentada a seguir

Definição 2. Suponha que queremos testar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ com $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \neq \emptyset$ e $\Theta_1 \neq \emptyset$. O TRVG rejeita H_0 para os pontos \mathbf{x} do espaço amostral que pertencem à região crítica

$$RC = \left\{ x = \frac{\sup_{\theta \in \Theta_1} L(\theta; x)}{\sup_{\theta \in \Theta_0} L(\theta; x)} \geq c \right\}$$

que também pode ser reescrita como:

$$RC = \left\{ \lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta; x)}{\sup_{\theta \in \Theta} L(\theta; x)} \leq c \right\}$$

Este teste Clássico também pode ser usado para hipóteses precisas e a regra de decisão é baseada na distribuição assintótica da estatística

$$-2 \log \lambda(\mathbf{x}) = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} \right),$$

onde $L(\theta; \mathbf{x})$ é a função de verossimilhança dos dados amostrais. A distribuição assintótica é qui-quadrado com ν graus de liberdade (χ_ν), e ν é dado pela dimensão de $\Theta \setminus \Theta_0$. O cálculo do p-value (p) é dado pela seguinte probabilidade, $p = Pr\{\chi_\nu > -2 \log \lambda(\mathbf{x})\}$.

2 Medidas de Evidência no Problema de Behrens-Fisher

Nesta seção apresentaremos a construção das medidas de evidência Clássica e Bayesiana, com base em amostras observadas das populações X e Y , respectivamente, $\mathbf{x} = \{x_1, \dots, x_n\}$ e $\mathbf{y} = \{y_1, \dots, y_m\}$. Para obtermos a região crítica do TRVG utilizamos a estatística

$$\lambda(\mathbf{x}, \mathbf{y}) = \left(\frac{\hat{\sigma}_x^2}{\hat{\sigma}_{x_0}^2} \right)^{\frac{n}{2}} \left(\frac{\hat{\sigma}_y^2}{\hat{\sigma}_{y_0}^2} \right)^{\frac{m}{2}},$$

com $\hat{\sigma}_x^2$ e $\hat{\sigma}_y^2$ representando, respectivamente, os estimadores de máxima verossimilhança (EMV) de σ_x^2 e σ_y^2 sob Θ , e $\hat{\sigma}_{x_0}^2$ e $\hat{\sigma}_{y_0}^2$ os EMV sob Θ_0 . Estes estimadores são dados por

$$\hat{\sigma}_x^2 = \frac{(n-1)S_x^2}{n}, \quad \hat{\sigma}_y^2 = \frac{(m-1)S_y^2}{m},$$

$$\hat{\sigma}_{x_0}^2 = \hat{\sigma}_x^2 + (\hat{\mu}_0 - \bar{X})^2, \quad \hat{\sigma}_{y_0}^2 = \hat{\sigma}_y^2 + (\hat{\mu}_0 - \bar{Y})^2$$

com \bar{X} e \bar{Y} representando as médias amostrais, S_x^2 e S_y^2 representando as variâncias amostrais, e

$$\hat{\mu}_0 = \frac{n\bar{X}\hat{\sigma}_{y_0}^2 + m\bar{Y}\hat{\sigma}_{x_0}^2}{n\hat{\sigma}_{y_0}^2 + m\hat{\sigma}_{x_0}^2}$$

Para a solução dos estimadores sob Θ_0 , $\hat{\mu}_0$, $\hat{\sigma}_{x_0}^2$ e $\hat{\sigma}_{y_0}^2$, será utilizado um procedimento numérico.

Para implementação do cálculo do FBST no problema de *Behrens-Fisher*, iremos considerar a priori imprópria $\pi(\theta) = \frac{c}{\sigma_x^2 \sigma_y^2}$ sugerida em Madruga(2002), e obtemos a densidade posteriori

$$\pi(\theta|\mathbf{x}, \mathbf{y}) = \pi_x(\mu_x, \sigma_x|\mathbf{x}) \times \pi_y(\mu_y, \sigma_y|\mathbf{y})$$

Sendo que $\pi_x(\mu_x, \sigma_x|\mathbf{x})$ e $\pi_y(\mu_y, \sigma_y|\mathbf{y})$ representam a f.d.p de uma distribuição normal-inversa-gama (NIG),

$$(\mu_x, \sigma_x|\mathbf{x}) \sim NIG(\hat{\sigma}_x^2, n-2, \bar{X}, 1/n)$$

e

$$(\mu_y, \sigma_y|\mathbf{y}) \sim NIG(\hat{\sigma}_y^2, m-2, \bar{Y}, 1/m).$$

Mas, $\pi_x(\mu_x, \sigma_x|\mathbf{x})$ e $\pi_y(\mu_y, \sigma_y|\mathbf{y})$ podem ser fatoradas como $\pi_x(\mu_x, \sigma_x|\mathbf{x}) = \pi_x(\mu_x|\sigma_x, \mathbf{x}) \times \pi_x(\sigma_x|\mathbf{x})$ e $\pi_y(\mu_y, \sigma_y|\mathbf{y}) = \pi_y(\mu_y|\sigma_y, \mathbf{y}) \times \pi_y(\sigma_y|\mathbf{y})$, com $\pi_x(\mu_x|\sigma_x, \mathbf{x})$ e $\pi_y(\mu_y|\sigma_y, \mathbf{y})$ representando distribuições normais com parâmetros $(\bar{X}, \sigma_x^2/n)$ e $(\bar{Y}, \sigma_y^2/m)$, respectivamente, e $\pi_x(\sigma_x|\mathbf{x})$ e $\pi_y(\sigma_y|\mathbf{y})$ representando o inverso de distribuições Gama. A partir destes modelos será gerada uma amostra aleatória de tamanho M do parâmetro de interesse θ , que serão usadas para aproximar a medida de evidência pelo método de Monte Carlo,

$$EV(\Theta_0; \mathbf{x}, \mathbf{y}) \approx 1 - \frac{1}{M} \sum_{j=1}^M h(\theta_j) \quad (1)$$

com

$$h(\theta) = 1(\theta \in T(\mathbf{x}, \mathbf{y})) = \begin{cases} 1 & \text{se } \theta \in T(\mathbf{x}, \mathbf{y}) \\ 0 & \text{se } \theta \notin T(\mathbf{x}, \mathbf{y}) \end{cases}$$

e

$$T(\mathbf{x}, \mathbf{y}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}, \mathbf{y}) > \pi(\hat{\theta}_0|\mathbf{x}, \mathbf{y})\},$$

onde $\hat{\theta}_0$ é o vetor de parâmetros que maximiza a posteriori sob Θ_0 .

3 Resultados

Apresentaremos nesta seção os resultados do nosso trabalho que foram obtidos a partir do software MATLAB (The Mathworks, Inc. [1999]). Resultados empíricos preliminares em alguns testes conhecidos (Madruga (2002)) indicam que a EV ($\Theta_0; \mathbf{x}$) pode ser obtida como a função de distribuição acumulada de um modelo de probabilidades Beta com parâmetros a e b , calculada no p-value correspondente, ou seja,

$$EV(\Theta_0, \mathbf{x}, \mathbf{y}) = \int_0^p \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1} du. \quad (2)$$

A tabela 1 apresenta as estimativas de a e b no ajuste da relação (1), para amostras de tamanho $n = m = 20$ geradas com diferentes valores da diferença $|\sigma_x^2 - \sigma_y^2|$. Em cada diferença foram calculados 21 pares de valores $(p, EV(\Theta_0, \mathbf{x}, \mathbf{y}))$, considerando as seguintes diferenças para as médias populacionais $|\mu_x - \mu_y| \in \{0, 0.1, 0.2, 0.3, \dots, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$. Por exemplo, para $|\sigma_x^2 - \sigma_y^2| = 1$, apresentamos na Tabela 2 os pares $(p, EV(\Theta_0, \mathbf{x}, \mathbf{y}))$, e na Figura 1 o gráfico de dispersão desses pares.

$ \sigma_x^2 - \sigma_y^2 $	\hat{a}	\hat{b}
0	0,3113	12,3754
1	0,4280	9,8500
2	0,4453	7,6504
3	0,4525	6,8372
4	0,4468	6,2379
5	0,4404	5,7341
6	0,4413	5,7785
7	0,4435	5,5687
8	0,4435	5,5687
9	0,4462	5,6640
10	0,4844	6,2369
11	0,4472	5,7232

Tabela 1: Estimativas de a e b segundo a diferença $|\sigma_x^2 - \sigma_y^2|$

$ \mu_x - \mu_y $	$EV(\Theta_0; \mathbf{x}, \mathbf{y})$	p
0.0	0.9990	0.7196
0.1	0.9360	0.1276
0.2	0.6060	0.0064
0.3	0.2850	0.0005
0.4	0.3360	0.0056
0.5	0.7220	0.0118
0.6	0.0540	0.0000
0.7	0.8570	0.0996
0.8	0.3520	0.0025
0.9	0.8960	0.1493
1.0	0.0260	0.0000
1.1	0.0160	0.0000
1.2	0.0330	0.0000
1.3	0.1280	0.0001
1.4	0.0000	0.0000
1.5	0.1110	0.0002
1.6	0.0030	0.0000
1.7	0.0000	0.0000
1.8	0.0000	0.0000
1.9	0.0000	0.0000
2.0	0.0000	0.0000

Tabela 2: Medidas de Evidência para $|\sigma_x^2 - \sigma_y^2| = 1$ e diferentes valores da diferença $|\mu_x - \mu_y|$

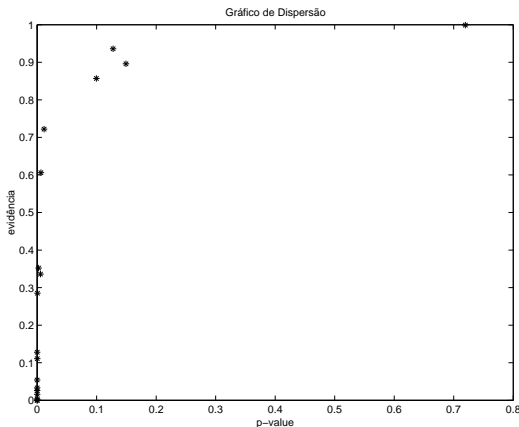


Figura 1: Gráfico de dispersão entre as medidas de evidência e o p-value da Tabela 2

Os resultados apresentados na Tabela 1 sugerem que as estimativas \hat{a} e \hat{b} estão sofrendo influência da diferença entre as variâncias, enquanto esta diferença é pequena, pois à medida que aumentamos a diferença entre as variâncias os resultados nos sugerem uma estabilidade das estimativas \hat{a} e \hat{b} .

Referências

- [1] A.O'Hagan, Kendall's Advanced Theory of Statistics 2B: Bayesian Inference.London: Edward Arnold, 1994.
- [2] C. A. de B. Pereira and J. Stern, Evidence and Credibility: a full Bayesian test of precise hypothesis, *Entropy*, 1 (1999) 99-110.
- [3] E. L. Lehmann, Testing Statistical Hypotheses, Wiley, New York, 1959.
- [4] G.E.P. Box e G.C.Tiao, Bayesian Inference in Statistical Analysis. Addison-wesley, Reading, MA, 1973.
- [5] H. Bolfarine e M. C. Sandoval, "Introdução à Inferência Estatística", Sociedade Brasileira de Matemática, Rio de Janeiro, 2001.
- [6] H. Jeffreys, "Theory of Probability", University Press, Oxford, 1961.
- [7] J. O. Berger e T. Selke, Testing a Point Null Hypothesis: The Irreconcilability of P values and Evidence, *Journal of the American Statistical Association*, 82 (1987) 112-139.
- [8] J. O. Berger e M. Delampady, Testing Precise Hypotheses, *Statistical Science*, 2 (1987) 317-352.
- [9] J.S.Mehta e R. Srinivasan, On the Behrens-Fisher problem, *Biometrika*, 57, 649-655, 1970.
- [10] K. A. Brownlee, Statistical theory and methodology in science and engineering. Wiley, New York, 1984.
- [11] M. R. Madruga, L. G. Esteves e S. Wechsler, On the Bayesianity of Pereira-Stern tests, *Test*, 10 (2001) 291-299.
- [12] M. R. Madruga, Teste de Significância: uma proposta genuinamente Bayesiana. 2002. 77 f. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.
- [13] The Mathworks, Inc. (1999). *MATLAB - The Language of Technical Computing*. Version 5.3.0.10183(R11).