

Relação Beta-Funcional entre o P-value e a Medida de Evidência Bayesiana na Comparação de duas Populações Normais

E.L.O. FERNANDES¹, L.O.O. DA COSTA², M.R. MADRUGA³, H.R. TAVARES⁴,
Programa de Pós-Graduação em Matemática e Estatística, CCEN / UFPA, Campus
do Guamá, 66075-110 Belém, PA, Brasil.

Resumo. [10] apresentou um procedimento para testar hipóteses precisas (hipóteses com medida de Lebesgue nula), que baseia-se no cálculo da probabilidade posterior da Região HPD (Highest Posterior Density) tangente ao conjunto que define a hipótese nula, definindo o complementar dessa probabilidade como uma medida de evidência Bayesiana em favor da hipótese precisa. Neste trabalho, utilizamos esta medida para testar a igualdade das médias de duas populações normais e, através de resultados de simulação, estabelecemos uma relação funcional entre esta medida e o p-value clássico do TRVG (Teste da Razão de Verossimilhanças Generalizada). Resultados empíricos preliminares indicam que a medida Bayesiana pode ser obtida pela função de distribuição acumulada do modelo Beta, calculada no p-value correspondente.

1. Introdução

O problema estatístico de *Teste de Hipóteses* consiste em rejeitar ou não uma hipótese H_0 sobre o valor de um parâmetro desconhecido com base na informação trazida pela amostra. Na abordagem Clássica é bastante usada uma medida de evidência denominada *p-value*, cujo cálculo baseia-se na distribuição amostral da estatística do teste. Na abordagem Bayesiana, as medidas de evidência usuais para teste de hipóteses são o *Fator de Bayes* e a *Probabilidade Posterior de H_0* . Alguns autores, [1] e [2], entre outros, apresentaram e discutiram os conflitos entre o *p-value* e as medidas de evidência Bayesianas, alertando para o fato de que em algumas situações o *p-value* pode não ser uma boa medida de evidência para uma hipótese estatística precisa (hipóteses com medida de Lebesgue nula). Com isso, surgiram novas medidas de evidência para teste de hipóteses precisas. Neste trabalho será apresentada a medida de evidência Bayesiana proposta por [10], cujo procedimento é denominado **Full Bayesian Significance Test (FBST)**. O objetivo é comparar, através de resultados de simulações, o *p-value* da abordagem Clássica, usando o

¹aluno de Mestrado do PPGME

²aluno de Mestrado do PPGME

³madruga@ufpa.br

⁴colaborador na produção deste trabalho

Teste da Razão de Verossimilhanças Generalizada (TRVG) ([3]) e a medida proposta por [10], sugerindo uma possível relação funcional entre essas duas medidas para testar a igualdade das médias de duas populações normais. Resultados empíricos preliminares indicam que a medida Bayesiana pode ser obtida pela função de distribuição acumulada do modelo Beta, calculada no *p-value* correspondente. Porém, como os parâmetros da distribuição Beta têm variado nos diferentes problemas estudados, outro objetivo é verificar se essa variação depende do tamanho amostral e da variância comum e conhecida das duas populações normais.

2. O Procedimento FBST

A medida de evidência em favor de uma hipótese precisa, introduzida por [10] é definida como

Definição 2.1. *Considere um modelo estatístico paramétrico, isto é, uma quintupla $(\mathcal{X}, \mathcal{A}, \mathbf{F}, \Theta, \pi)$, onde \mathcal{X} é um espaço amostral, \mathcal{A} é uma sigma-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 Θ . Suponha que um subconjunto Θ_0 de Θ tendo medida de Lebesgue nula 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 $EV(\Theta_0, \mathbf{x}) = 1 - P[\theta \in T(\mathbf{x})|\mathbf{x}]$.*

Como podemos ver da Definição 2.1, a medida de evidência de Pereira-Stern considera, em favor de uma hipótese precisa, 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 [10], 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; por outro lado, um valor “pequeno” da $EV(\Theta_0, \mathbf{x})$ indica que Θ_0 está em uma região do espaço paramétrico de baixa probabilidade posterior, portanto, os dados nos levariam a desacreditar da hipótese nula. Uma vantagem deste procedimento é que ele contorna a dificuldade de tratar com uma hipótese precisa (detalhes em [6]), pois não há necessidade de introduzir uma probabilidade positiva a priori como no teste Bayesiano padrão ([5]). [10] defendem que o uso de $EV(\Theta_0, \mathbf{x})$ para avaliar a evidência trazida pelos dados para Θ_0 é um procedimento “Bayesiano”, uma vez que apenas a densidade posterior está envolvida. [7] 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 (ver [11]). Assim, o teste de Pereira-Stern (FBST) consiste em

- **Rejeitar H_0 se $EV(\Theta_0, \mathbf{x}) \leq K$,**
- **Aceitar H_0 se $EV(\Theta_0, \mathbf{x}) > K$,**

onde K é um ponto crítico cujo valor depende da função de perda escolhida. Por exemplo, [7] consideram $\mathbf{D} = \{Aceitar H_0 (d_0), Rejeitar H_0 (d_1)\}$ o espaço de decisões e definem a função de perda $L : \mathbf{D} \times \Theta \rightarrow \mathbb{R}^+$, dada por $L(Rejeitar H_0, \theta) = a[1 - \mathbf{1}(\theta \in T(\mathbf{x}))]$ e $L(Aceitar H_0, \theta) = b + c\mathbf{1}(\theta \in T(\mathbf{x}))$, com $a, b, c > 0$, sendo $\mathbf{1}(\theta \in T(\mathbf{x}))$ a função indicadora do conjunto $T(\mathbf{x})$. Eles mostram que para essa função de perda tem-se $K = \frac{b+c}{a+c}$. A propriedade de Invariância da $EV(\Theta_0, \mathbf{x})$ é obtida através de uma transformação de coordenadas na região $T(\mathbf{x})$, e é apresentada em [9], com base no *Princípio da Surpresa Mínima* sugerido em [4].

3. Teste da Razão de Verossimilhanças Generalizada

Seja Θ o espaço paramétrico, ou seja, o conjunto de valores possíveis para θ . 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 $\Theta_1 \neq \emptyset$. Então o TRVG tem região crítica, ou seja, a região dos pontos do espaço amostral, $\mathbf{x} \in \mathcal{X}$, que levam à rejeição de H_0 , dada por

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

que também pode ser reescrita como

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

com $L(\theta; \mathbf{x})$ representando a *função de verossimilhança*.

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

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

onde os graus de liberdade da distribuição Qui-quadrado são dados por $\nu = \dim(\Theta \setminus \Theta_0)$, ou seja, é a diferença entre o número de parâmetros desconhecidos em Θ e o número de parâmetros desconhecidos sob Θ_0 . Logo, o *p-value* (p) é dado por

$$p = P(\chi_\nu^2 \geq -2 \log \lambda(\mathbf{x})).$$

4. Relação Funcional entre $EV(\Theta_0, \mathbf{x})$ e *p-value*

Resultados empíricos preliminares em alguns testes conhecidos ([8]) 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,

$$EV(\Theta_0, \mathbf{x}) = \int_0^p \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1} du. \quad (4.1)$$

As estimativas de a e b na relação (4.1) são obtidas através do ajuste de modelos de regressão não-lineares, e têm variado nos testes já estudados e baseados em simulações([8]).

Evidentemente, o estatístico Bayesiano não precisa relacionar soluções Bayesianas com soluções Clássicas, porém no caso dos testes de hipóteses a solução Clássica (p-value) possui uma regra de decisão bem definida, que é a comparação do p-value com um nível de significância pré-estabelecido pelo pesquisador, enquanto que a regra de decisão da medida de evidência proposta por [10] depende de uma função de perda escolhida, o que acarreta considerável dificuldade. Uma forma de contornar esta dificuldade é estudando uma possível relação funcional entre estas duas soluções, que leve à construção de uma regra de decisão que não contemple a necessidade de trabalhar com funções de perda.

5. Comparação das Médias de duas Normais

Com base em amostras obtidas das duas populações, X_1, \dots, X_n uma amostra aleatória de $X \sim N(\mu_1, \sigma^2)$, Y_1, \dots, Y_m uma amostra aleatória de $Y \sim N(\mu_2, \sigma^2)$, e admitindo σ^2 conhecida, queremos testar as hipóteses $H_0 : \mu_1 = \mu_2$ contra $H_1 : \mu_1 \neq \mu_2$, ou seja, tem-se

$$\Theta = \{(\mu_1, \mu_2) : \mu_1 \in \mathbb{R} \text{ e } \mu_2 \in \mathbb{R}\}$$

$$\Theta_0 = \{(\mu_1, \mu_2) \in \Theta : \mu_1 = \mu_2\}.$$

A função de verossimilhança dos dados observados (\mathbf{x}, \mathbf{y}) , com $\mathbf{x} = \{x_1, \dots, x_n\}$ e $\mathbf{y} = \{y_1, \dots, y_m\}$, é dada por

$$L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n+m}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [\delta_1 + \delta_2] \right\}, \quad (5.1)$$

com $\delta_1 = (n-1)S_x^2 + n(\mu_1 - \bar{x})^2$ e $\delta_2 = (m-1)S_y^2 + m(\mu_2 - \bar{y})^2$, sendo $\bar{x} = (1/n) \sum_{i=1}^n x_i$ e $\bar{y} = (1/m) \sum_{j=1}^m y_j$ as médias amostrais, $S_x^2 = (1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})^2$ e $S_y^2 = (1/(m-1)) \sum_{j=1}^m (y_j - \bar{y})^2$ as variâncias amostrais.

5.1. Construção da $EV(\Theta_0; \mathbf{x}, \mathbf{y})$ usando o procedimento FBST

Considerando a priori imprópria usual para $\boldsymbol{\theta} = (\mu_1, \mu_2)$, $\pi(\boldsymbol{\theta}) = c$, temos que a densidade posterior é proporcional ao produto da priori pela função de verossimilhança dos dados observados, ou seja,

$$\pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} [n(\mu_1 - \bar{x})^2 + m(\mu_2 - \bar{y})^2] \right\}, \quad (5.2)$$

que pode ser fatorada como o produto das densidades posteriores de μ_1 e μ_2 , $\pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) = \pi_1(\mu_1|\mathbf{x}) \times \pi_2(\mu_2|\mathbf{y})$, com as funções $\pi_i(\cdot)$, $i = 1, 2$, representando, respectivamente, as funções densidade de probabilidade dos seguintes modelos de probabilidade: $\mu_1|\mathbf{x} \sim N(\bar{x}; \sigma^2/n)$ e $\mu_2|\mathbf{y} \sim N(\bar{y}; \sigma^2/m)$. Para o cálculo da

$EV(\Theta_0; \mathbf{x}, \mathbf{y})$, primeiramente maximizamos a distribuição posterior $\pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$ sob a hipótese nula, obtendo

$$\pi(\hat{\mu}|\mathbf{x}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} [n(\hat{\mu} - \bar{x})^2 + m(\hat{\mu} - \bar{y})^2] \right\},$$

com $\hat{\mu} = (n\bar{x} + m\bar{y})/(n + m)$. Em seguida, resolvemos a seguinte integral

$$I = \int_{T(\mathbf{x}, \mathbf{y})} \pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) d\boldsymbol{\theta}$$

com $T(\mathbf{x}, \mathbf{y}) = \{\boldsymbol{\theta} \in \Theta : \pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) > \pi(\hat{\mu}|\mathbf{x}, \mathbf{y})\}$. A integral acima pode ser expressa como a esperança posterior da função indicadora $h(\boldsymbol{\theta}) = 1(\boldsymbol{\theta} \in T(\mathbf{x}, \mathbf{y}))$, ou seja,

$$I = E[h(\boldsymbol{\theta})|\mathbf{x}, \mathbf{y}] = \int_{\Theta} 1(\boldsymbol{\theta} \in T(\mathbf{x}, \mathbf{y})) \pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) d\boldsymbol{\theta}, \tag{5.3}$$

que pode ser aproximada pelo Método de Monte Carlo. Para isso, simulamos uma amostra aleatória de tamanho M , $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M)$, da distribuição em (5.2) e aproximamos a integral (5.3) por

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M h(\boldsymbol{\theta}_j).$$

Logo, a medida de evidência do procedimento FBST é aproximada por

$$EV(\Theta_0; \mathbf{x}, \mathbf{y}) \approx 1 - \frac{1}{M} \sum_{j=1}^M h(\boldsymbol{\theta}_j).$$

5.2. Construção do *p-value* usando o TRVG

A estatística do TRVG é dada por

$$-2 \log \lambda(\mathbf{x}, \mathbf{y}) = \frac{1}{\sigma^2} [n(\hat{\mu} - \bar{x})^2 + m(\hat{\mu} - \bar{y})^2],$$

e usando sua distribuição assintótica χ_1^2 , obtemos o *p-value*

$$p = \int_{-2 \log \lambda(\mathbf{x}, \mathbf{y})}^{+\infty} \frac{(1/2)^{1/2}}{\Gamma(1/2)} u^{\frac{1}{2}-1} e^{-\frac{u}{2}} du.$$

6. Resultados

A fim de detectar se há influência dos tamanhos amostrais (n e m) e da variância comum conhecida (σ^2) nas estimativas dos parâmetros "a" e "b" da relação beta-funcional (relação 4.1), planejou-se um estudo de simulação em três etapas. Na etapa I, fixados os valores de n , m e σ^2 , foram geradas 20 pares de amostras de tamanhos n e m das distribuições normais $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$, respectivamente, para $\mu_1 = 0$ e 20 diferentes valores de μ_2 . Para cada par de amostra gerada, foram obtidos os valores de p e $EV(\Theta_0; \mathbf{x}, \mathbf{y})$ para testar as hipóteses

$H_0 : \mu_1 = \mu_2$ contra $H_1 : \mu_1 \neq \mu_2$ e, em seguida, ajustou-se a relação 4.1 aos pares de valores $(p, EV(\Theta_0; \mathbf{x}, \mathbf{y}))$, obtendo-se as estimativas de "a" e "b". Para exemplificar, a Tabela 1 apresenta o p -value (p) e a $EV(\Theta_0; \mathbf{x}, \mathbf{y})$ para $n = m = 30$, $\sigma^2 = 9$, e os diferentes valores de μ_2 . A Figura 1 apresenta o gráfico de dispersão dos pares $(p, EV(\Theta_0; \mathbf{x}, \mathbf{y}))$ apresentados na Tabela 1. Ajustando o modelo (4.1) aos pontos da Figura 1, obtivemos as estimativas $\hat{a} = 0,8423$ e $\hat{b} = 2,0468$.

$ \mu_1 - \mu_2 $	$EV(\Theta_0; \mathbf{x}, \mathbf{y})$	p
0,0	0,9642	0,7835
0,1	0,7653	0,4608
0,2	0,9820	0,8585
0,3	0,5939	0,3052
0,4	0,7234	0,4164
0,5	0,7064	0,4055
0,6	0,9966	0,9294
0,7	0,9465	0,7348
0,8	0,1328	0,0434
0,9	0,0437	0,0124
1,0	0,9857	0,8696
1,2	0,0010	0,0002
1,4	0,1896	0,0683
1,6	0,2517	0,0981
2,0	0,0077	0,0018
2,5	0,0038	0,0010
3,0	0,0000	0,0000
3,5	0,0001	0,0000
4,0	0,0000	0,0000
4,5	0,0000	0,0000

Tabela 1: Medidas de Evidência para o teste de igualdade das médias de duas populações normais ($n = m = 30$, $\sigma^2 = 9$, $\mu_1 = 0$)

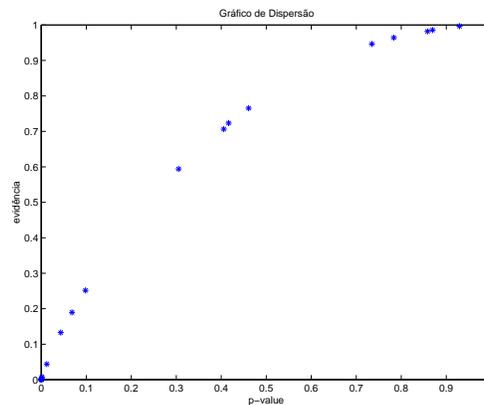


Figura 1: Gráfico de dispersão entre as medidas de evidência da Tabela 1

A etapa II consistiu em repetir a etapa I para $\sigma^2 = 9$ e diferentes valores comuns de n e m , a fim de detectar a influência do tamanho amostral nas estimativas de a e b . A Tabela 2 apresenta estas estimativas, sugerindo que as mesmas estabilizam-se em torno de $\hat{a} = 0,84$ e $\hat{b} = 2,04$. A etapa III consistiu em repetir a etapa I para

$n = m = 30$ e diferentes valores de σ^2 , a fim de detectar a influência da variância nas estimativas de a e b . A Tabela 3 apresenta estas estimativas, sugerindo também que as mesmas estabilizam-se em torno de $\hat{a} = 0,84$ e $\hat{b} = 2,04$.

Valores para n e m	a	b
05	0,84	2,04
10	0,85	2,05
15	0,83	2,03
20	0,84	2,05
25	0,85	2,07
30	0,84	2,05
35	0,83	2,01
40	0,83	2,03
45	0,83	2,03
50	0,85	2,04
55	0,83	2,00
60	0,84	2,03
65	0,84	2,04
70	0,84	2,04
100	0,83	2,01
200	0,85	2,07
300	0,85	2,07
600	0,84	2,02

Tabela 2: Estimativas de a e b para diferentes valores de n e m e $\sigma^2 = 9$

Valores para σ^2	a	b
01	0,83	2,01
03	0,83	2,03
05	0,85	2,07
07	0,84	2,05
09	0,84	2,05
11	0,85	2,06
13	0,84	2,06
15	0,84	2,05
17	0,85	2,05
19	0,84	2,04
21	0,84	2,05
23	0,85	2,05
25	0,85	2,05
27	0,85	2,06
29	0,85	2,05
70	0,84	2,04
90	0,84	2,06
200	0,84	2,05

Tabela 3: Estimativas de a e b para $n = m = 30$ e diferentes valores de σ^2

7. Conclusões

Os resultados apresentados nas Tabelas 2 e 3 sugerem que as estimativas de a e b , no ajuste da relação (4.1), não dependem das dimensões amostrais e nem da variância comum conhecida, no exemplo estudado. Em um trabalho futuro,

pretende-se verificar a possibilidade de estabelecer uma distribuição assintótica para a medida de evidência $EV(\Theta_0; \mathbf{x}, \mathbf{y})$, uma vez que a mesma pode ser considerada uma estatística de teste, na visão Clássica dos testes de hipóteses. Esta distribuição assintótica, assim como acontece no TRVG, pode sugerir uma possível dependência entre as estimativas de a e b e o número de parâmetros desconhecidos envolvidos no estudo.

Abstract. [10] presented a procedure in order to test precise hypotheses (i.e., hypotheses having null Lebesgue measure), that are based on posterior probabilities of the HPD (Highest Posterior Density) tangent to the set that defines the null hypothesis, defining the complement of this probability as a measurement of Bayesian evidence in favour of a precise hypothesis. In this work this measurement is used to test equality of the means in two populations distributed normally, and through simulation, a functional relation is established between this measurement and the classic p-value of the Generalized Likelihood Ratio Test. Early empirical results indicate that the Bayesian measurement can be obtained from the Beta cumulative distribution function, calculated for the corresponding p-value.

Referências

- [1] J.O. Berger, T. Selke, Testing a point null hypothesis: the irreconcilability of p values and evidence, *Journal of the American Statistical Association*, **82** (1987), 112-139.
- [2] J.O. Berger, M. Delampady, Testing precise hypotheses, *Statistical Science*, **2** (1987), 317-352.
- [3] H. Bolfarine, M.C. Sandoval, “Introdução à Inferência Estatística”, Sociedade Brasileira de Matemática, Rio de Janeiro, 2001.
- [4] M. Evans, Bayesian inference procedures derived via the concept of relative surprise. *Communications in Statistics*, **26** (1997), 1125-1143.
- [5] H. Jeffreys, “Theory of Probability”, University Press, Oxford, 1961.
- [6] D.V. Lindley, A Statistical Paradox, *Biometrika*, **44** (1957), 187-192.
- [7] M.R. Madruga, L.G. Esteves, S. Wechsler, On the Bayesianity of Pereira-Stern tests, *Test*, **10** (2001), 291-299.
- [8] M.R. Madruga, “Teste de Significância: Uma Proposta Genuinamente Bayesiana”, Tese de Doutorado, IME, USP, São Paulo, SP, 2002.
- [9] M.R. Madruga, C.A. de B. Pereira, J. Stern, Bayesian evidence test for precise hypotheses, *Journal of Statistical Planning and Inference*, **117** (2003), 185-198.
- [10] C.A. de B. Pereira, J. Stern, Evidence and credibility: a full bayesian test of precise hypothesis, *Entropy*, **1** (1999), 99-110.
- [11] H. Rubin, A weak system of axioms for 'rational' behaviour and the non-separability of utility from prior, *Statistics and Decisions*, **5** (1987), 47-58.