

EXERCICIO PROGRAMA 6 - MAP2212 2024

Victor Rocha Cardoso Cruz

11223757

Larissa Aparecida Marques Pimenta Santos

12558620

1 Enunciado

- Escreva um programa para computar o e -valor padronizado da hipótese de Hardy-Weinberg no modelo Dirichlet-Trinomial, para os diversos conjuntos de observações, \mathbf{X} , descritos em C. A. B. Pereira, J. M. Stern (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. Entropy, 1, 69-80 [1].
- Em cada caso, compute a função verdade do modelo, \mathbf{W} , usando o código desenvolvido para o 5º EP.
- Use uma rotina de otimização numérica disponível em seu ambiente de programação para encontrar $\theta^* = \arg \max_{\theta \in H} s(\theta)$, e então compute $s^* = s(\theta^*)$, $ev(H | X) = W(s^*)$, e $sev(H | X)$.
- Para quais conjuntos de observações, \mathbf{X} , você aceitaria/rejeitaria (ou estaria indeciso sobre) \mathbf{H} ? Por quê?
- Compute os e -valores (padronizados) usando as prioris plana ($y = 1$) ou nula ($y = 0$). Suas conclusões mudariam? Por quê?

2 Objetivos

O principal objetivo deste exercício é calcular o e -valor (ou valor epistêmico) para a Hipótese de Hardy-Weinberg, no modelo Dirichlet trinomial. O conceito de e -valor foi inicialmente introduzido por Pereira e Stern em 1999, num artigo para o periódico *entropy* [1]. Tal conceito se apresenta como uma alternativa ao clássico modelo de aceitação/rejeição da hipótese baseada no p -valor, devido a algumas limitações conhecidas deste último [1]. A tabela 2 do artigo mencionado é utilizada como insumo para os cálculos realizados, bem como parte do código baseia-se no algoritmo implementado nos Exercícios Programa 04 e 05. Por fim, com base nos cálculos realizados e no e -valor encontrado, a hipótese é, ou não, rejeitada.

3 Funcionamento do Programa

Inicialmente, são carregados na memória todos os valores para os vetores X e Y (tabela 2 do artigo), que serão utilizados para as futuras operações. São selecionados, inicialmente, os primeiros vetores X e Y da lista. Geram-se 3.850.000 amostras utilizando o gerador Dirichlet. Essas amostras são utilizadas para calcular valores da função potencial.

Os valores da função potencial são inseridos num vetor, chamado *vetor valores*, que é então ordenado e separado em 2.500 grupos. São colocados num segundo vetor, chamado *vetor fronteira*, os maiores valores de cada grupo. Então calcula-se o maior valor que a função potencial da hipótese pode atingir. Esse cálculo é realizado por uma função específica da biblioteca de funções SciPy (*minimize_scalar*, do subgrupo *optimize*).

Esse maior valor é normalizado e a função verdade, $W(v)$, é calculada. A aproximação de $W(v)$ é dada pela divisão entre o valor da variável e o valor total de pontos gerados. O resultado desse cálculo é o e -valor (e). O próximo passo é calcular $\bar{e} = 1 - e$. O \bar{e} é utilizado para o cálculo da função QQ definida pelo nos slides da aula 06, que representa o *sev*. O e -valor padronizado (*sev*) é dado por:

$$sev = 1 - s\bar{e}v$$

Finalmente a hipótese é ou não rejeitada, comparando-se o *sev* obtido com o valor limiar fixado em 0,05. Se for menor, a hipótese é rejeitada, caso contrário, é aceita. O mesmo processo é repetido para os demais valores da tabela 2, e os resultados vão sendo apresentados ao usuário à medida que são calculados.

Há 2 valores que foram desconsiderados da tabela, com $x_3 = 0$, pois ambos desencadeiam erros nos cálculos. Os demais 34 casos foram calculados para dois vetores $y : [0, 0, 0]$ e $[1, 1, 1]$. Em 4 desses casos, houve divergência de resultados quanto à rejeição/não rejeição da hipótese para os dois vetores y :

$x = [1, 13, 6]$, $x = [1, 12, 7]$, $x = [5, 13, 2]$ e $y = [0, 0, 0]$: rejeita hipótese

$x = [1, 13, 6]$, $x = [1, 12, 7]$, $x = [5, 13, 2]$ e $y = [1, 1, 1]$: não rejeita hipótese

Para o caso:

$x = [1, 1, 18]$ e $y = [0, 0, 0]$: não rejeita hipótese

$x = [1, 1, 18]$ e $y = [1, 1, 1]$: rejeita hipótese

Para os demais 30 casos, para ambos vetores y , os resultados de rejeição/não rejeição foram iguais. De forma resumida:

$y = [0, 0, 0]$: 23 não rejeições

$y = [1, 1, 1]$: 25 não rejeições

4 Conclusão

A rejeição/não rejeição da hipótese nula, H_0 , baseou-se no critério do nível de significância $\alpha = 0,05$, comumente utilizado em testes de hipóteses, valor esse que limita a probabilidade de erro tipo I em 5%. Note-se que outros valores de α poderiam ser utilizados, mas nesse trabalho optou-se em aplicar o valor normalmente adotado.

O vetor $y = [1, 1, 1]$, correspondente à *priori plana*, trouxe resultados muito parecidos com o vetor $y = [0, 0, 0]$, que corresponde a não ter observações *a priori*. Dos 34 casos considerados a partir da tabela 2, em 30 deles o vetor y não influenciou na rejeição/não rejeição da hipótese nula. Em 4 casos houve divergência.

Percebe-se que existe uma clara influência nos resultados, em cerca de 10% dos casos. Porém, nesse formato, é um fator que não altera de maneira significativa os resultados. Uma possível continuação desse trabalho seria explorar outros vetores y , para dessa forma extrair exatamente a influência que ele causa nos resultados.

Referências

- [1] PEREIRA, C. A. B. e STERN, J. M., *Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses*, Entropy, 1999.