Adaptive significance level test using weighted likelihoods for comparison of proportions.

Andrés F. Flórez Rivera.
Luís G. Esteves.
Instituto de Matemática e Estatística - Universidade de São Paulo.

ISBA - 2018- World-Meeting.

Abstract

Statistical hypothesis testing is a widely used mathematical tool in various fields of knowledge. Nevertheless, it is a controversial topic due to the misuse of canonical (fixed) values of significance level which frequently leads to unreasonable decisions or statistical paradoxes, such as Lindleys and Bartletts paradoxes.

In this work, we apply the procedure based on adaptive significance levels proposed by [1][4], to the problem of equality of proportions on contingence tables. Weighted likelihood methodology is used for such procedure, instead of usual maximum likelihood values over the hypotheses[2][3]. The new procedure overcomes difficulties with fixed significance levels and it is in accordance with the likelihood principle.

Statistical Model

.

Let $L(\theta|x)$ be the likelihood function and Θ_H, Θ_A be a partition of parametric space Θ, such that, $\Theta_H \bigcup \Theta_A = \Theta$ and $\Theta_H \bigcap \Theta_A = \emptyset$. The hypotheses to be considered are:

$H: \theta \in \Theta_H$ where each probabiliy will be described by $\mathbb{P}(\theta \in \Theta_H) = \xi_H$
$A: \theta \in \Theta_A$ and $\mathbb{P}(\theta \in \Theta_A) = \xi_A$.

A prior θ has probability density function $g(\theta)$ on the entire parameter space Θ, and posterior density function $g(\theta|x)$.

Bayesian Test

The proposed test is based on the predictive distribution of the data under each hypothesis, H and A, where:

$$f_H(x) = \frac{\int g(\theta)L(\theta|x)d\Theta_H}{\int g(\theta)d\Theta_H} \quad \text{and} \quad f_A(d) = \frac{\int g(\theta)L(\theta|x)d\Theta_A}{\int g(\theta)d\Theta_A} \tag{1}$$

and $d\Theta_H$ ($d\Theta_A$) denotes that the integral line (surface) is being taken for each hypothesis. Let δ^* be a test such that:

$$\delta^* = \begin{cases} 1 & \frac{f_H(x)}{f_A(x)} < b/a \\ 0 & o.c \end{cases} \tag{2}$$

and accept any if $af_H(x) = bf_A(x)$, with $a > 0$ and $b > 0$. So, how was showed by [4][3], δ^* is an optimal test that minimize the error sum, i.e:

$$a \times \alpha(\delta^*) + b \times \beta(\delta^*) \leq a \times \alpha(\delta) + b \times \beta(\delta)$$

The type I and II error probabilities α_{δ^*} and $\beta_{\delta^*}$is given by:

$$\alpha_{\delta^*} = \sum_{\substack{x\in\mathfrak{X}:\\ BF(x)\leq b/a}} \left(\frac{\int g(\theta)L(\theta|x)d\Theta_H}{\int g(\theta)d\Theta_H}\right) = \sum_{\substack{x\in\mathfrak{X}:\\ BF(x)\leq b/a}} f_H(x) \qquad \beta_{\delta^*} = \sum_{\substack{x\in\mathfrak{X}:\\ BF(x)> b/a}} \left(\frac{\int g(\theta)L(\theta|x)d\Theta_A}{\int g(\theta)d\Theta_A}\right) = \sum_{\substack{x\in\mathfrak{X}:\\ BF(x)> b/a}} f_A(x)$$

If the value x_0 was observed, so the $P-valor(x_0)$ index is:

$$P-valor(x_0) = \sum_{\substack{x\in\mathfrak{X}:\\ f_H(x)>BF(x_0)\times f_A(x)}} f_H(x)$$

Binomial Case

We are going to use (1) to test the following hypotheses:

$$H: \theta_1 = \theta_2 = \ldots = \theta_m \quad \text{and} \quad A: \theta_i \neq \theta_j \ \text{ for } i \neq j \in 1, 2, \ldots m \tag{3}$$

each θ_i is the proportion of successes of m populations, where, $x_i \sim Bin(n_i, \theta_i)$ and $\theta_i \sim Beta(a_i, b_i)$. Thus:

$$L(\theta|\mathbf{X}) = \prod_{i=1}^{m} \binom{n_i}{x_i} \theta_i^{x_i}(1-\theta_i)^{n_i-x_i} \qquad g(\theta) = \prod_{i=1}^{m} \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta_i^{a_i-1}(1-\theta_i)^{b_i-1}$$

So, the predictive distributions of the data $\mathbf{X} = (x_1, x_2, \ldots, x_m)$ are given by:

$$f_H(\mathbf{X}) = \frac{\prod_{i=1}^m \binom{n_i}{x_i}\Gamma(B)\Gamma(C)\Gamma(\sum_{i=1}^m (a_i+b_i)-2(m-1))}{\Gamma(B+C)\Gamma(\sum_{i=1}^m a_i-(m-1))\Gamma(\sum_{i=1}^m b_i-(m-1))} \qquad f_A(\mathbf{X}) = \prod_{i=1}^m \binom{n_i}{x_i} \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \prod_{i=1}^m \frac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)}$$

Hence, the posterior odds in favor of H is given by:

$$BF(\mathbf{X}) = \frac{\Gamma(A)\Gamma(B)\Gamma(\sum_{i=1}^m (a_i+b_i)-2(m-1))}{\Gamma(A+B)\Gamma(\sum_{i=1}^m a_i-(m-1))\Gamma(\sum_{i=1}^m b_i-(m-1))\prod_{i=1}^m D_i} \quad \text{Where} \quad D_i = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \frac{\Gamma(a_i+x_i)\Gamma(n_i+b_i-x_i)}{\Gamma(n_i+b_i+a_i)}$$

Multinomial Case

We are going to use one more time (1) to test the following hypotheses:

$$H: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad \text{and} \quad A: \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \tag{4}$$

each $\boldsymbol{\theta}_i = (\theta_{1i}, \theta_{2i} \ldots \theta_{ki})$ is the proportion of successes of 2 populations, where, k is the levels number, $\boldsymbol{X}_i \sim Multinomial(n_i, \boldsymbol{\theta}_i)$ and $\boldsymbol{\theta}_i \sim Dir(\boldsymbol{\alpha}_i)$, for $i = 1, 2.$. Thus:

$$L(\theta|\mathbf{X}) = \prod_{j=1}^2 \left\{ \frac{\Gamma(\sum_{i=1}^k x_{ij}+1)}{\prod_{i=1}^k \Gamma(x_{ij}+1)} \prod_{i=1}^k \theta_{ij}^{x_{ij}} \right\} \qquad g(\theta) = \prod_{j=1}^2 \left\{ \frac{\Gamma(\sum_{i=1}^k \alpha_{ij})}{\prod_{i=1}^k \Gamma(\alpha_{ij})} \prod_{i=1}^k \theta_{ij}^{\alpha_{ij}-1} \right\}$$

So, the predictive distributions of the data $\mathbf{X} = (x_1, x_2, \ldots, x_m)$ are given by:

$$f_H(\mathbf{X}) = \prod_{j=1}^2 \left\{ \frac{\Gamma(\sum_{i=1}^k x_{ij}+1)}{\prod_{i=1}^k \Gamma(x_{ij}+1)} \right\} \frac{\prod_{i=1}^k \Gamma(\alpha_i^*)}{\Gamma(\sum_{i=1}^k \alpha_i^*)} \frac{\Gamma(\sum_{i=1}^k \alpha_{i1}+\alpha_{i2}-1)}{\prod_{i=1}^k \Gamma(\alpha_{i1}+\alpha_{i2}-1)} \qquad f_A(\mathbf{X}) = \prod_{j=1}^2 \left\{ \frac{\Gamma(\sum_{i=1}^k \alpha_{ij})\Gamma(\sum_{i=1}^k x_{ij}+1)}{\prod_{i=1}^k \Gamma(\alpha_{ij})\prod_{i=1}^k \Gamma(x_{ij}+1)} \right\} \frac{\prod_{i=1}^k \Gamma(\alpha_{i1}+x_{i1})\prod_{i=1}^k \Gamma(\alpha_{i2}+x_{i2})}{\Gamma(\sum_{i=1}^k \alpha_{i1}+x_{i1})\Gamma(\sum_{i=1}^k \alpha_{i2}+x_{i2})}$$

Hence, the posterior odds in favor of H are:

$$BF(\mathbf{X}) = \frac{\frac{\prod_{i=1}^k \Gamma(\alpha_i^*)}{\Gamma(\sum_{i=1}^k \alpha_i^*)} \frac{\Gamma(\sum_{i=1}^k \alpha_{i1}+\alpha_{i2}-1)}{\prod_{i=1}^k \Gamma(\alpha_{i1}+\alpha_{i2}-1)}}{\prod_{j=1}^2 \left\{ \frac{\Gamma(\sum_{i=1}^k \alpha_{ij})}{\prod_{i=1}^k \Gamma(\alpha_{ij})} \right\} \frac{\prod_{i=1}^k \Gamma(\alpha_{i1}+x_{i1})\prod_{i=1}^k \Gamma(\alpha_{i2}+x_{i2})}{\Gamma(\sum_{i=1}^k \alpha_{i1}+x_{i1})\Gamma(\sum_{i=1}^k \alpha_{i2}+x_{i2})}} \tag{5}$$

where $\alpha_i^* = (\alpha_{i1} + \alpha_{i2} + x_{i1} + x_{i2} - 1)$.

Examples

Binomial Case

We are interested in testing if the proportion of success is the same on three different populations, hence, the hypothesis is $H: \theta_1 = \theta_2 = \theta_3$. Let that the assessed priors are the uniform Beta distributions, i.e, $a_i = b_i = 1$ for $i = 1, 2, 3.$, the results were:

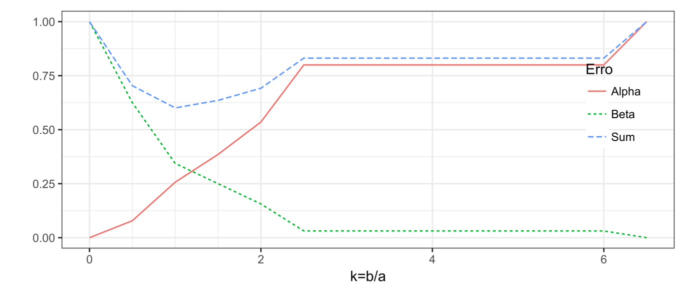

Figure 1: α, β and its sum for differents of k with $n_1 = n_2 = n_3 = 3$.

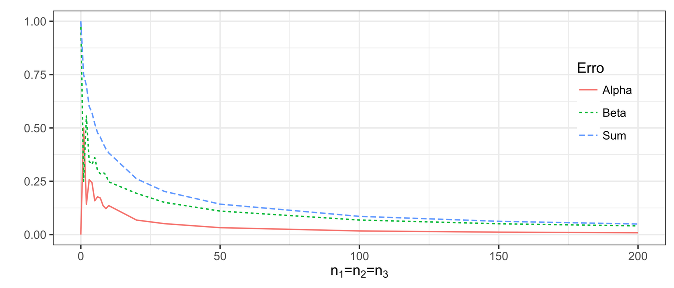

Figure 2: Optimal α, β and its sum for differents sample size, balanced case.

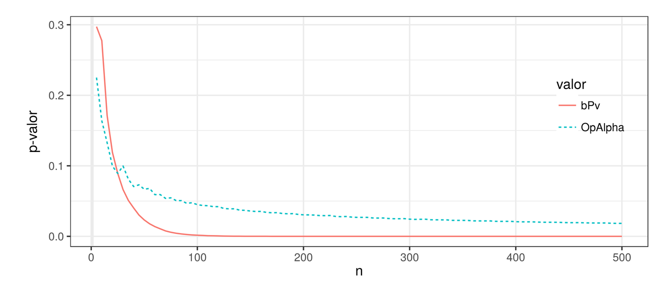

Figure 3: P-valor for different sample size by BF test.

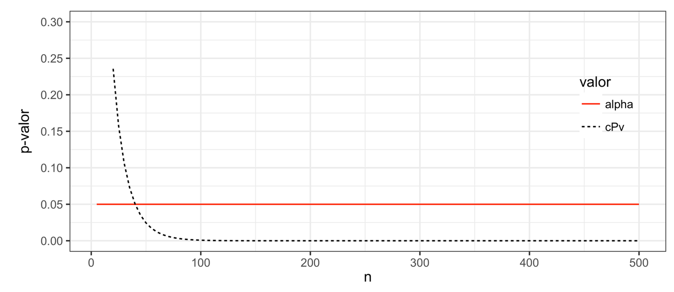

Figure 4: p-valor for differents sample size by Pearsons Chi- squared.

Multinomial Case

We are interested in testing if the proportion of success is the same on two different populations each one with 3 levels, hence, the hypothesis is $H: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Let that the assessed priors are the uniform Dirichlet distributions, i.e, $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{1}.$, so, we have that:

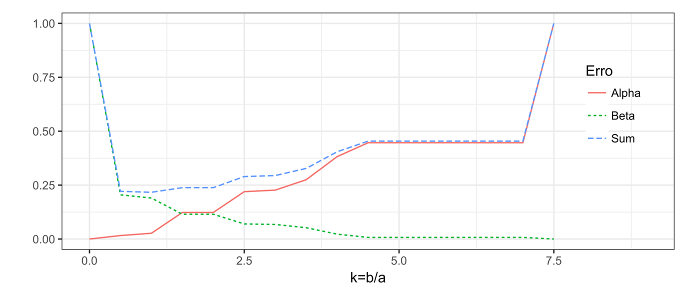

Figure 5: α, β and its sum for differents of k with $n_1 = n_2 = 3$.

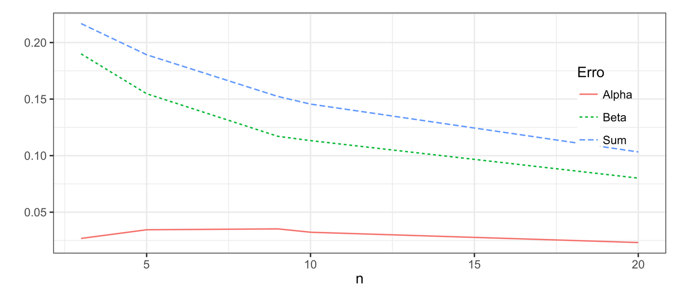

Figure 6: Optimal α, β and its sum for differents sample size, balanced case.

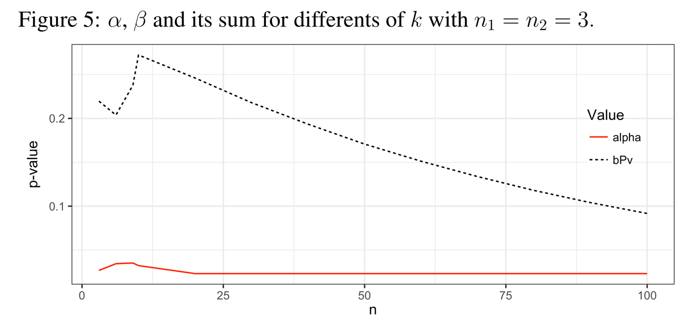

Figure 7: P-valor for different sample size by BF test.

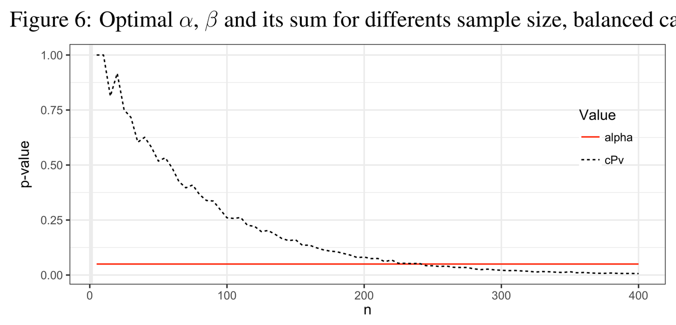

Figure 8: p-valor for differents sample size by Pearsons Chi- squared.

Remarks

- Weighted likelihoods are a straightforward tool when we want to test sharp hypotheses, for both the Binomial and Multinomial case.
- Optimal significance level decreasing as the sample size increases. This is helpful when needed to estimate the sample size, for instance, if in the Binomial case we want to have a level of significance of 0.025 the sample size minimum should be 250.

Acknowledgments

The authors gratefully acknowledges financial support from graduate program in statistics of the university of São Paulo CAPES/PROEX.

References

[1] M DeGroot. *Probability and Statistics*. Adson-Wesley: Boston, MA, USA, 1986.

[2] T. Irony and C Pereira. Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas IME-USP*, 2(1):2746, 1995.

[3] C. Pereira, E. Nakano, V. Fossaluza, L. Esteves, M. Gannon, and A. Polpo. Hypothesis tests for bernoulli experiments: Ordering the sample space by bayes factors and using adaptive significance levels for decisions. *Entropy*, 19:696, 2017.

[4] L. Pericchi and C. Pereira. Adaptive significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Brazilian Journal of Probability and Statistics*, 30(1):70–90, 2016.