A comparison of two Poisson means using weighted likelihoods and the NNP principle.

Andrés F. Flórez Rivera.
Luís G. Esteves.
Instituto de Matemática e Estatística - Universidade de São Paulo.
Lima-Perú - COBAL-2019.

Abstract

The problem of statistical hypothesis testing consist of a decision problem in which the objective is to choose a statistical hypothesis between two hypotheses. Each of these hypotheses defines a specific subset of the parameter space. Usually, some of these subsets do not have the same dimension and this makes it difficult to compare them. Among several approach to this problem, we discuss a solution based in weighted likelihoods which are defined under the subset of the parameter space specific to each hypothesis. This idea was initially proposed by (Irony and Pereira, 1995).

This proposal aims to avoid the "Statistical Paradox" showed by Lindley (1957) and the misuse of canonical values of significance. Adaptive significance levels are applied to compare the means of two Poisson distributions with unknown parameters. For this propose, the analysis is carried out by conditioning on a suitable statistic in order to make use of Berger and Wolpert's principle of the non-informative nuisance parameter J. and R. (1988).

Statistical Model

Let $x \in \mathfrak{X} \subset \Re^s$ denote the data obtained as result of an experiment and $\lambda \in \Omega \subset \Re^k$ the unknown parameter that identifies the function $f(x|\lambda)$, where X and Ω are the sample space and parameter space respectively. Suppose that we can partition Ω as $\Omega = \Omega_1 \cup \Omega_2$, where, $\Omega_1 \bigcap \Omega_2 = \emptyset$. Hence, we are concerned in testing the following hypotheses:

$$\begin{aligned} H &: \lambda \in \Omega_0 \\ A &: \lambda \in \Omega_0^c \end{aligned} \tag{1}$$

In addition, let $L(\boldsymbol{\lambda}|\boldsymbol{x})$ be the likelihood function for $\boldsymbol{\lambda}$ generated by x. A prior $\boldsymbol{\lambda}$ has probability density function $\pi(\boldsymbol{\lambda})$ on the entire parameter space Ω.

Weighted Likelihoods

The proposed test is based on the prior predictive functions $f_H(\boldsymbol{x})$ and $f_A(\boldsymbol{x})$ as evidence measures of the data $\boldsymbol{x}$ under each hypothesis, H and A, where:

$$f_H(\boldsymbol{x}) = E\left(L(\boldsymbol{\lambda}|\boldsymbol{x})|\boldsymbol{\lambda} \in \Omega_0\right) \quad \text{and} \quad f_A(\boldsymbol{x}) = E\left(L(\boldsymbol{\lambda}|\boldsymbol{x})|\boldsymbol{\lambda} \in \Omega_0^c\right) \tag{2}$$

Thus, the Bayes factor in favor of H is given by $BF(\boldsymbol{x}) = \frac{f_H(\boldsymbol{x})}{f_A(\boldsymbol{x})}$. Let δ^* be a test such that:

$$\delta^* = \begin{cases} 1 & BF(\boldsymbol{x}) < b/a \\ 0 & otherwise, \end{cases} \tag{3}$$

with $a, b > 0$. The type I and II error probabilities α_{δ^*} and β_{δ^*} are **Adaptive Significance levels** given by:

$$\alpha_{\delta(x)^*} = \sum_{\substack{\boldsymbol{x} \in \mathfrak{X} \\ BF(\boldsymbol{x}) \leq b/a}} f_H(\boldsymbol{x}) \quad \text{and} \quad \beta_{\delta(x)^*} = \sum_{\substack{\boldsymbol{x} \in \mathfrak{X} \\ BF(\boldsymbol{x}) > b/a}} f_A(\boldsymbol{x}) \tag{4}$$

In addition, if the value x_0 is observed, the *P-value*(x_0) index is given by

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

NNP Principle

Suppose that $\lambda = (\lambda_1, \lambda_2)$ and the likelihood function $L(\boldsymbol{\lambda}|\boldsymbol{x})$ can be factorized as

$$L(\boldsymbol{\lambda}|\boldsymbol{x}) = L^1(\lambda_1|\boldsymbol{x}) \times L^2(\lambda_2|\boldsymbol{x}) \tag{5}$$

Definition 1 J. and R. (1988): Suppose E_1 is an experiment such that (5) is satisfied. Let E_2 be the "thought" experiment in which, in addition to $\boldsymbol{x}$, λ_2 is observed, in this case the parameter is λ_1. Then λ_2 is a noninformative nuisance parameter if $Ev(E_2, (x, \lambda_2))$ is independent of λ_2.

The NNP principle states that if E_1 is as in (5) and λ_2 is as noninformative nuisance parameter as in definition 1, then

$$Ev(E_1, x) = Ev(E_2, (x, \lambda_2))$$

that is, if one were to reach the identical conclusion for every λ_2, were λ_2 known, then that the same conclusion should be reached even if λ_2 is unknown (J. and R. (1988)).

Poisson means comparison

We shall test equality of two Poisson means. For this, let $\mathbf{x} = (x_1, x_2)$ be the independent observations that follow a Poisson distribution with parameters λ_1 and λ_2 respectively, where $\lambda_1 \in \Omega_1$, $\lambda_2 \in \Omega_2$ and $\Omega = \Omega_1 \times \Omega_2$. The hypotheses of equality of means can be formulated as in (1): $\Omega_0 = \{(\lambda_1, \lambda_2) : \lambda_1 = \lambda_2\}$.

The likelihood function for this problem is given by

$$L(\boldsymbol{\lambda}|\boldsymbol{x}) = \frac{\lambda_1^{x_1}\lambda_2^{x_2}}{x_1!\, x_2!} e^{-\lambda_1} e^{-\lambda_2}$$

The prior distribution for $\boldsymbol{\lambda}$ can be expressed by

$$\pi(\boldsymbol{\lambda}) = \frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \lambda_1^{a-1} \lambda_2^{c-1} e^{-\lambda_1 b} e^{-\lambda_2 d}$$

In this way, the prior predictive functions (2) under each hypothesis is given by

$$\begin{aligned} f_H(\boldsymbol{x}) &= \int_\Omega L(\boldsymbol{\lambda}|\boldsymbol{x}) \pi(\boldsymbol{\lambda\lambda} \in \Omega_0) d\mathbb{P}_H(\boldsymbol{\lambda}) \\ &= \frac{1}{x_1!} \frac{1}{x_2!} \frac{\Gamma(x_1+x_2+a+c-1)(b+d)^{(a+c-1)}}{(b+d+2)^{(x_1+x_2+a+c-1)}\Gamma(a+c-1)} \end{aligned}$$

$$\begin{aligned} f_A(\boldsymbol{x}) &= \int_\Omega L(\boldsymbol{\lambda}|\boldsymbol{x}) \pi(\boldsymbol{\lambda}|\boldsymbol{\lambda} \in \Omega_0^c) d\mathbb{P}_A(\boldsymbol{\lambda}) \\ &= \frac{b^a}{\Gamma(a)} \frac{d^c}{\Gamma(c)} \frac{1}{x_1!} \frac{1}{x_2!} \frac{\Gamma(x_1+a)}{(1+b)^{(x_1+a)}} \frac{\Gamma(x_2+c)}{(1+d)^{(x_2+c)}} \end{aligned}$$

When $a = b = c = d = 1$, the posterior odds in favor of H are given by:

$$BF(\mathbf{x}) = 2 \times \left(\frac{1}{2}\right)^{(x_1+x_2)} \binom{x_1+x_2}{x_1}$$

The test introduced by Irony and Pereira (1995) and Pericchi and Pereira (2016) meets the NNP principle presented before. To apply the NNP principle in this case, we need conditioning in some statistics such that the likelihood can be factored as in (5). Then,

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2|\boldsymbol{\lambda}) &= P(X_1 = x_1|X_1 + X_2 = x_1 + x_2, \boldsymbol{\lambda}) P(X_1 + X_2 = x_1 + x_2|\boldsymbol{\lambda}) \\ &= \binom{n}{x_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x_2} \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^{x_1+x_2}}{(x_1+x_2)!} \end{aligned}$$

Consequently,

$$L(\boldsymbol{\lambda}|\boldsymbol{x}) = \binom{x_1+x_2}{x_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x_2} \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^{x_1+x_2}}{(x_1+x_2)!} \tag{6}$$

Consider now the re-parametrization

$$\gamma_1 = \frac{\lambda_1}{\lambda_1+\lambda_2}; \qquad \gamma_2 = \lambda_1 + \lambda_2; \tag{7}$$

with, $\Omega \to \Lambda = [0,1] \times R_+$, where $\gamma_1 \in \Lambda_1$, $\gamma_2 \in \Lambda_2$ and $\Lambda = \Lambda_1 \times \Lambda_2$. This new parametrization allows the likelihood (6) to be written as

$$L(\boldsymbol{\gamma}|\boldsymbol{x}) = L^1(\boldsymbol{\gamma}_1|\boldsymbol{x}) L^2(\boldsymbol{\gamma}_2|\boldsymbol{x}) \tag{8}$$

Hence, from the NNP principle, it is not necessary to take into account $L^2(\boldsymbol{\gamma}_2|\boldsymbol{x})$ to make inferences concerning γ_1. Thus, suppose that the prior distribution for γ_1 can be given by

$$\pi(\gamma_1) = \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} \gamma_1^{a-1} (1-\gamma_1)^{c-1}$$

Note that with the re-parametrization the hypotheses (1) are rewritten as

$$\begin{aligned} H' &: \gamma_1 \in \Lambda_0 \\ A' &: \gamma_1 \in \Lambda_0^c \end{aligned} \tag{9}$$

with $\Lambda_0 = \{1/2\} \times \Lambda_2$. Under H', $\pi(\gamma_1)$ has degenerate distribution at $1/2$, hence, $\pi(\gamma_1|\gamma_1 = 1/2) = 1$. The prior predictive functions under each hypothesis is given by

$$\begin{aligned} f_{H'}(\boldsymbol{x}) &= L^x(\gamma_1|\boldsymbol{x}, \gamma_1 = 1/2) \\ &= \binom{x_1+x_2}{x_1} (1/2)^{x_1+x_2} \end{aligned}$$

$$\begin{aligned} f_{A'}(\boldsymbol{x}) &= \int_{\Lambda_1} L^1(\gamma_1|\boldsymbol{x}) \pi(\gamma_1|\gamma_1 \in \Lambda 1) d\mathbb{P}_A(\gamma_1) \\ &= \binom{x_1+x_2}{x_1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \gamma_1^{x_1+a-1} (1-\gamma_1)^{x_2+b-1} d\gamma_1 \end{aligned}$$

As a result, when $a = b = c = d = 1$, the posterior odds in favor of H are given by

$$BF(\boldsymbol{x}) = (x_1 + x_2 + 1) \left(\frac{1}{2}\right)^{(x_1+x_2)} \binom{x_1+x_2}{x_1} \tag{10}$$

The adaptive significance level $\boldsymbol{\alpha}$ and the *P-value* are computed for different sample sizes in the case when $L(\boldsymbol{\gamma}|\boldsymbol{x})$ is used and when the NNP principle is applied ($L^1(\gamma_1|\boldsymbol{x})$ is used):

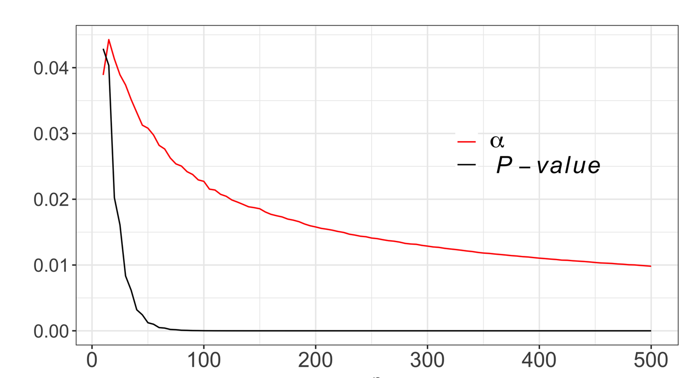

Figure 1: Adaptive $\boldsymbol{\alpha}$ and *P-value* for different sample sizes taking into account all parameters.

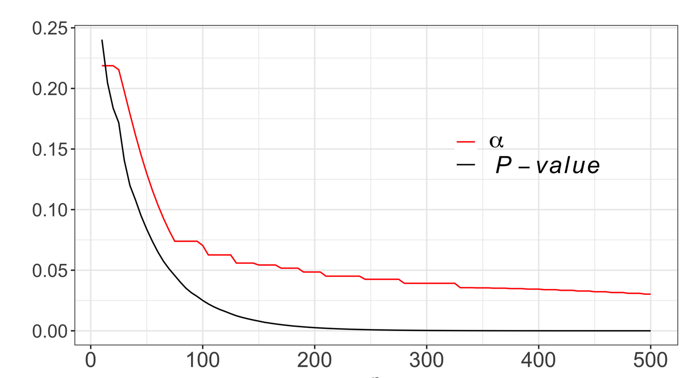

Figure 2: Adaptive $\boldsymbol{\alpha}$ and *P-value* for different sample sizes following the NNP principle.

From **Figure 1** and **2** the *P-value* is always smaller than α. Consequently, the conclusion will be the same in both cases.

Remarks

- To apply NNP principle to the procedures based on weighted likelihoods are make it easier to test "sharp" hypotheses: to compare Poisson means reduce to test a simple hypothesis.
- Optimal significance level decreasing as the sample size increases. This is helpful when there is interest to estimate the sample size. For instance, in the **Figure 1** we can see that to achieve a level of significance smaller than 0.02 the minimum sample size should be 120.

Acknowledgments

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

References

Irony and Pereira (1995) T. Irony and C Pereira. Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas IME-USP*, 2(1):27–46. From page

J. and R. (1988) Berger J. and Wolpert R. *The likelihood principle*. Hayward, CA: Institute of Mathematical Statistics. From page

Lindley (1957) D Lindley. A statistical paradox. *Biometrika*, 44:187–192. From page

Pericchi and Pereira (2016) 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. From page