II Encuentro Matemático del Caribe

Universidad Tecnológica de Bolívar & Universidad del Sinú Seccional Cartagena

Septiembre 09 - 12, 2020, Cartagena de Indias - Colombia

Adaptive α Significance Level for Linear Models

Tipo: Ponencia

Daiver Vélez, Luis Pericchi, María Pérez [*]

Resumen

We put forward an adaptive alpha that decreases as the information grows, for hypothesis tests in which nested linear models are compared. A less elaborate adaptation was already presented in Pérez and Pericchi (2014) for comparing general i.i.d. models. In this article we present refined versions to compare nested linear models. This calibration may be interpreted as a Bayes-non-Bayes compromise, and leads to statistical consistency, and most importantly, it is a step forward towards statistics that leads to reproducible scientific findings.

Palabras & frases claves: p-value calibration; Bayes factor; linear model; likelihood ratio; adaptive alpha; PBIC .

1. Introducción

The p-value associated with hypothesis testing is defined as the minimum significance level with which the null hypothesis would be rejected in favor of the alternative or a correct interpretation of the p-values is to visualize them as the probability of obtaining results as or more extreme as the observed result when the null hypothesis is true. In general, a significance level α=0.05 is considered, and then the decision rule would be: reject the null hypothesis in favor of the alternative hypothesis if the p-value is less than 0.05, or not to reject the null hypothesis in favor of the alternative hypothesis if the p-value is greater than 0.05.

[*]Universidad de Puerto Rico, Río Piedras, e-mail: daiver.velez@upr.edu

When the null hypothesis is rejected, the results are said to be statistically significant. But the p-values say nothing about the magnitudes of the effects. A non-significant p-value does not imply that the null hypothesis is true. If large enough samples are used, it is highly likely to reject a false null hypothesis. In this case, a non-significant p-value provides evidence in favor of the null hypothesis. But if large samples are not used, there is little likelihood of rejecting a false null hypothesis. In this case, a non-significant p-value does not provide evidence in favor of the null hypothesis. When the samples are not large enough, we have no way of knowing whether the non-significant p-values are due to a certain null hypothesis or to the unlikely probability of rejecting a false null hypothesis due to the unsuitable sample size (Sellke et al.).
Bayesian literature have been criticizing for several decades the implementation of hypothesis testing with fixed significance levels, and in particular the use of the scale p-value<0.05. That discussion was mostly regarded as a philosophical issue about the wrong interpretation of p-values as probabilities of the null hypothesis. However, the crisis of Fisher's scale of evidence exploded when scientific researchers, largely outside departments of Mathematics and Statistics, began reporting very low rates of reproducible scientific presumed findings. Many papers and opinions have been written on this subject, and we will mention just a few of them as in (Pericchi and Pérez).

- In 2005, John Ioannidis publish a paper in PLOS Medicine entitled "Why Most Published Research Findings Are False".
- Sir David Cox: "Statistics is also about science and p <0.05 is seen as the passport for publication, and most if not all statisticians would take a rather disapproving view of it to put it mildly".
- In 2015, the Basic and Applied Social Psychology Journal banned the use of significance testing, so p < 0.05 will not any more be a "passport for publication".
- In March 2016, the American Statistical Association publishes a Statement on Statistical Significance and p-Values, establishing some general principles for the use and interpretation of p-values, principles on which all Statisticians could agree.
- In September 2017, a group of more than 70 researchers in Statistics published a paper asking for "redefining statistical significance". They propose to change the default p-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.
- In March 2019, Valentin Amrhein, Sander Greenland and Blake McShane publish a paper in Nature 567, 305-307, about:"Scientists rise up against statistical significance".

It is then clear that obtaining a p-value lower than 0.05 does not open the doors for publication as before, and we Statisticians must provide alternatives

to Scientists.

Referencias

[1] Pérez, M. E. and L. R. Pericchi (2014). Changing statistical significance with the amount of information: The adaptive alfa significance level. Statistics and Probability Letters 85, 20-24.

[2] Sellke, T., M. J. Bayarri, and J. O. Berger (2001). Calibration of p values for testing precise null hypotheses. The American Statistician 55 (1), 62-71.