# ON THE STATISTICAL ASSESSMENT OF SMALL SAMPLE CLASSIFICATION

Q. Noirhomme[1], D. Lesenfants[1], F. Gomez[2], A. Soddu[3], A. Luxen[1], C. Phillips[1] and S. Laureys[1]

[1]Cyclotron Research Centre, University of Liège, Liège, Belgium
[2]Computer Science Department, Universidad Central de Colombia, Bogotá, Colombia
[3]Department of Physics & Astronomy, Western University, London Ontario, Canada

**Abstract**

Classifiers start to be used in medical application to infer diagnosis. Their results are assessed through either a binomial or a permutation test. Distributions built from classification of random data with cross-validation, did not follow the theoretical binomial distribution, showing that binomial test was not conservative enough. A permutation test is thus recommended.

Keyword(s): bioinformatics – biosignals

## 1. INTRODUCTION

Machine learning approaches are now commonly used to study brain functions [2] and have been proposed as a diagnostic tool for patients. The validity of the classification results is assessed through a statistical test. Some limitations have been highlighted in a recent study on patients with disorders of consciousness where a reanalysis of the data drew opposite conclusions [1]. In the present paper, we used simulated data to compare the properties of the binomial and permutation tests on small sample data. Both tests have been proposed in the literature to assess classification results [2].

## 2. SIMULATION

We generated random datasets with varying number of trials and features for a two-class problem. The features were randomly assigned 0 and 1. A label was also randomly assigned to each trial. We tested several cross-validation schemes with a “linear discriminant analysis” classifier. Every time the accuracy was above 0.5 we run a permutation test with 999 permutations.

## 3. RESULTS

We report here only the distribution of 10000 simulations with 100 trials, 40 features and 10x10-folds cross-validation scheme. These results are representative of the results obtained with the different sets of parameters and cross-validation schemes.

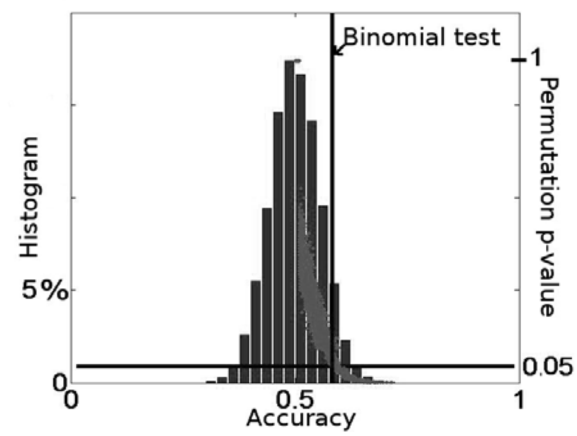


Figure 1. Histogram of the distribution of the classification accuracy (bars) and p-values from the permutation test (light grey stars)

The binomial test has a lower bound at 0.583 (resp. 0.616) for 100 trials (Fig. 1, vertical line) and $p<.05$ (resp. $p<.01$), leading to more than 7% (resp. 2%) of false positives. For the permutation test, 4.9% (resp. 0.92%) of all p-values were below 0.05 (resp. 0.01).

## 4. CONCLUSION

The cross-validation scheme has an influence on the independence of the data for small sample dataset. This influence biased the binomial test. The permutation test takes the cross-validation scheme into account and is not influenced. Therefore a permutation test is recommended especially when dealing with small sample sizes and non-independent cross-validation schemes.

**References**

[1] Goldfine AM. et al. Reanalysis of "Bedside detection of awareness in the vegetative state: a cohort study". *Lancet*, 381(9863):289-91, 2013.

[2] Pereira F. et al. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage,* 45 S199–S209, 2009.