Estimating log likelihood ratio test power using Kullback Leibler divergence on brain measurements

Nitay Alon[1] and Isaac Meilijson[1]

Abstract

We present an approach to estimate the power of a log likelihood ratio test (llrt) using Kullback Leibler divergence (KLD). Our method allows to estimate the number of samples which a log likelihood ratio test will yield over a 50% power. The method was applied on brain features measurements data and the findings are reported

1. INTRODUCTION

The human brain is one of the most complex system in the biological world. In recent years mixture models are becoming a major part in estimating the structure of the human brain. It is hypothesize that men and women brains are infect a mixture of base distribution. To test this hypothesis we preform a log-likelihood ratio test between the null, pure typed brain, and the alternative likelihood. In order to pre-estimate the hypothesis we've used a relation between the asymptotic distribution of the llrt under the alternative and the KLD to estimate the percentile of the Chi-Square distribution. Our findings showed that for some combinations of the mixture model parameters it is highly likely to reject the null hypothesis and for other combinations it is non likely.

2. Model and Data

2.1. Model

We model the composition of a human brain as a mixture model of two sex based distributions, that is, a men brain is a mixture model of both masculine and feminine features and a women is the a mixture as well. The genders are separated by the mixture probabilities as seen in the following equations:

$$f_{men}(x) = p * f_{masculine}(x) + (1-p) * f_{feminine}(x) \quad (1)$$

$$f_{women}(x) = q * f_{masculine}(x) + (1-q) * f_{feminine}(x) \quad (2)$$

Where $f_Y(y)$ is an exponential family member distribution.

2.1.1. Log likelihood ratio test.
We test the following hypothesis:

$$H_0 : p = (1-q) = 1 \quad (3)$$
$$H_1 : 0 < q < p < 1 \quad (4)$$

And the likelihood ratio between the genders can be written as:

$$\lambda = log\big(\frac{p * f_{masculine}(x) + (1-p) * f_{feminine}(x)}{q * f_{masculine}(x) + (1-q) * f_{feminine}(x)}\big) \quad (5)$$

As we can see, if the null hypothesis is true then the former likelihood ratio test is reduced to:

$$\lambda = log\big(\frac{f_{masculine}(x)}{f_{feminine}(x)}\big)$$

and since we've assumed that $f_X(x)$ is a continuous member of the exponential family this log likelihood ratio is a linear line with limits:

$$\lim_{x \to -\infty} log\big(\frac{f_{masculine}(x)}{f_{feminine}(x)}\big) = -\infty$$

$$\lim_{x \to \infty} log\big(\frac{f_{masculine}(x)}{f_{feminine}(x)}\big) = \infty$$

On the other hand, if the alternative is true then the log likelihood ratio is a sigmoid with the following limits:

$$\lim_{x \to -\infty} log\big(\frac{f_{men}(x)}{f_{women}(x)}\big) = \frac{1-p}{1-q}$$

$$\lim_{x \to \infty} log\big(\frac{f_{masculine}(x)}{f_{feminine}(x)}\big) = \frac{p}{q}$$

and we can see that this is true regardless to the underline distribution $f_X(x)$.

2.2. Data

The data was simulated using R software. We've created two simulations - one with Normal distribution

and one with Asymmetric Double Exponential distribution as the underling sex distribution. To simulate the data we've created a grid of the mixture probability $p \in \{0.51, \cdots, 1\}$ and the empirical group means $\bar{x}$ (since the data is normalized the men mean = -women mean). We then computed θ - the mean of the underling distribution. Using the set $\{theta, p\}$ we were able to compute the group variance and sampled from the mixture model.

3. KLD and power estimation

We use KLD to estimate the power of the test.

3.1. KLD

We use MC simulation to estimate the KLD between two distributions:

$$K\hat{L}D = \frac{1}{n} \sum_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)} \qquad (6)$$

and the data was sampled from $f_1(x)$.

3.2. Asymptotic distribution of llkr

To estimate the power of the test we've used the following relation between the asymptotic distribution of the lltr under the alternative and the KLD:

$$2 * \sum_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}|_{H_1} \geq n * KLD \qquad (7)$$

4. Findings

We report one the findings in the case of a normal underling distribution and in the case of a ADL distribution.

4.0.1. Normal distribution. In the case of Normal underling distribution we find that for most pairs $\{\bar{x}, p\}$ it is almost impossible to reject the null hypothesis (pure types) even if the data was generated from a mixture model (type 2 error) - thus implying that the log likelihood ratio test has weak power.

5. CONCLUSIONS

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

APPENDIX

ACKNOWLEDGMENT

References