

# dissertation\_\_writeup\_\_draft

*Alassane Ndour*

*23/08/2019*

## 1 Introduction and Objectives

### 1.1 background to the problem

- Importance of growth models
- Noisy environment - Systematic error
- Classification using Bayes factor - e.g. See for a dataset classified as linear which model seems to fit (frequentist) better which will be selected by BF and by how much while varying noise

### 1.2 Reasons for the choice of project

- Application in biology and economics
- Contribution to literature as unexplored method - combination of Bayes factor and Harris paper
- Study of noise to signal ratio in classification and different types of noise

### 1.3 Identification of the project's beneficiaries

- Commercial partner (probably not but might get data from them)
- Literature as an empirical analysis of Bayesian classification of growth using different models

### 1.4 Objectives and metrics

- A Classification framework which should include :
  - A classification between different models
  - The “certainty” of classification - TBD how we can quantify this
  - An estimation of the parameters of the model - with the “certainty” of estimation
  - An identification of the systematic error

### 1.5 Broad methods and how they answer goal

- Curve fitting :
  - Fit a linear and a logistic and classify depending on the error. See how as you increase the variance of the error, the classification changes (currently doing)
- Bayesian approach :
  - Estimate the distribution of the parameters (we should get the “certainty” from here) of a Bayesian linear regression, sigmoid function and then add the algorithm set by Harris (his calculation was for a sigmoid. Might have to do it for a linear regression) and compare the models using Bayes factor.
  - Does the model that fits the most correspond to the correct functional form?

- See if as you change error the systematic error is caught by the Harris algo and how the model selection varies
- Compare the two approaches : how do they compare ? In terms of classification error rate for instance
- Furthermore, there have been interesting developments in combining Bayesian methods and cross validation as they are not mutually exclusive methods and can contribute to robust estimates. Such works include Bürkner et al. (2019) where the authors aim to improve upon leave-future-out cross-validation (LFO-CV) - an adaptation of leave-one-out cross-validation (LOO-CV) to timeseries - to reduce computation time.

## 2 Context (Literature)

- Ed Harris
- LOO-CV
- Bayesian books
- Bürkner et al. (2019)
- Claeskens and Lid Hjort (2010) (Model selection)
- 

## 3 Data

- Synthetic data:
  - Two sets of data one linear and one sigmoid (describe the parameters of sigmoid)
  - Split datasets into ratio of error to signal. buckets currently range from 0 to 1 with intervals of 0.1 (variance of the error)

## 4 Methods

Here each subsection can correspond to one interest point of the experiment (e.g. adding drift, taking only the certain quartiles of the signal...). Each subsection can be divided into curve fitting led approaches and bayesian approaches - With ultimately a pipeline in mind as end goal. The data is divided into bucket so that the overall dataset looks like this : (dataset contains the numpy arrays of the data in question, variance\_bucket is the variance of the error going from 0 to 1 - when 1 the error is std normal, drift\_line is BOOL describes if a drift line was added to the signal and functional form is "logistic" or "linear" )

```
head(dplyr::tibble(c("dataset", "variance_bucket", "drift_line", "functional_form")))
```

```
## # A tibble: 4 x 1
##   `c("dataset", "variance_bucket", "drift_line", "functional_form")`
##   <chr>
## 1 dataset
## 2 variance_bucket
## 3 drift_line
## 4 functional_form
```

#### **4.1 Frequentist - curve fitting**

- find how we would tackle systematic error in this case - TBD

#### **4.2 Bayesian model selection**

- find how we would tackle systematic error in this case - Ed Harris ##

### **5 Results**

### **6 Discussion**

### **7 Evaluation, Reflections, and Conclusions**