SYDE 372 - Lab 2

Nikhil Malhotra – 20368862

Miguel de Arcos – 20400557

# Introduction

The purpose of this lab is to analyze statistical model estimation and classifier aggregation. Specifically 1D and 2D data sets will be analyzed for parametric, nonparametric and sequential estimation. For the 1D a Gaussian, Exponential and Uniform distributions will be investigated utilizing the Maximum Likelihood estimation technique. The Parzen estimation technique will be utilized for the non-parametric case for Gaussian only in both a 1D case and the 2D case. The sequential estimator will only be used for the 2D data sets.

# Model Estimation 1-D Case

The following data sets were utilized for this section:

* variable a - a bunch of Gaussian samples,
* variable b - a bunch of Exponential samples

## Parametric Estimation:

### Gaussian:

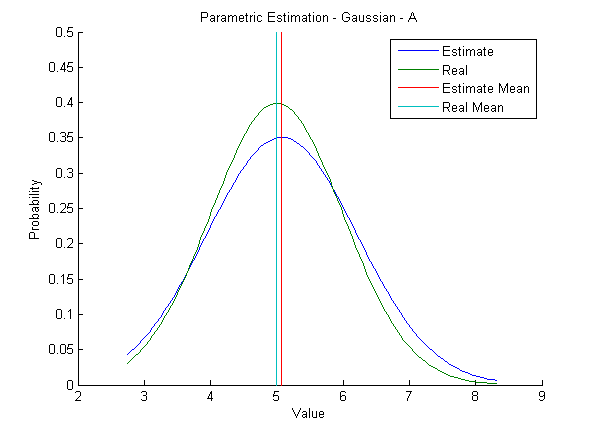


Figure 1: Gaussian ML Estimation on Normal Data

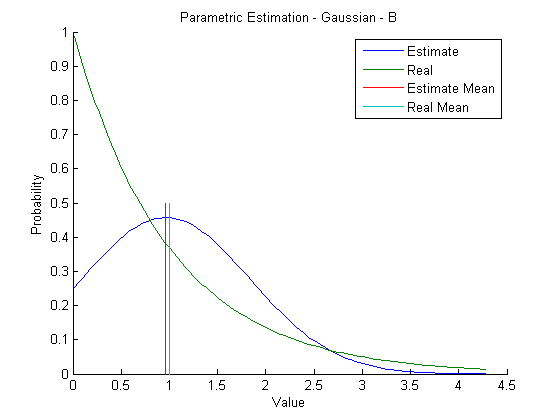


Figure 2: Gaussian ML Estimation on Exponential Data

Observing the graphs above, the closest estimated pdf is for data set a. This is obvious as data set is normally distributed unlike data set b, which is exponentially distributed.

### Exponential:

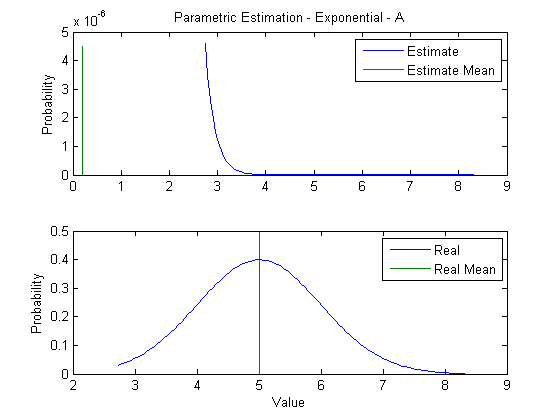


Figure : Exponential ML Estimation on Normal Data

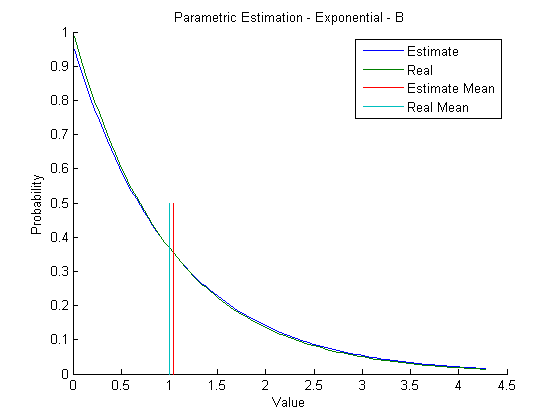


Figure : Exponential ML Estimation on Exponential Data

Observing the graphs above, the closest estimated pdf is for data set b. This is obvious as data set is exponentially distributed unlike data set a, which is normally distributed.

### Uniform:

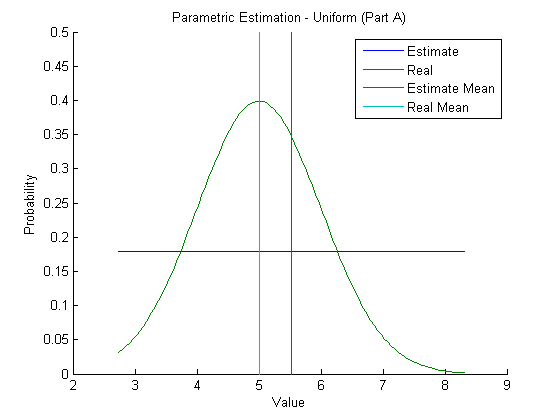


Figure : Uniform ML Estimation on Normal Data

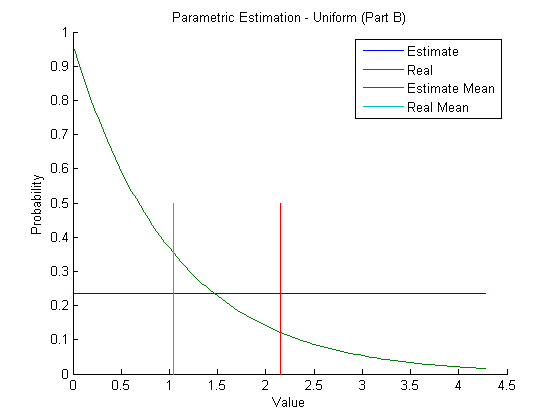


Figure : Uniform ML Estimation on Exponential Data

Observing the graphs above, the closest estimated pdf is for data set a. Although it is closer, it is ill suited for data set a.

## Non-Parametric Estimation:

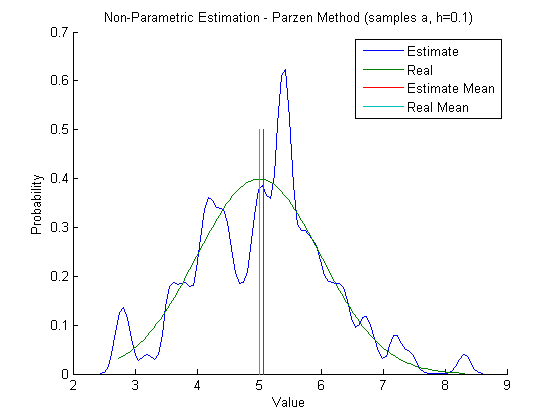


Figure : Parzen Estimation on Normal Data, h =0.1

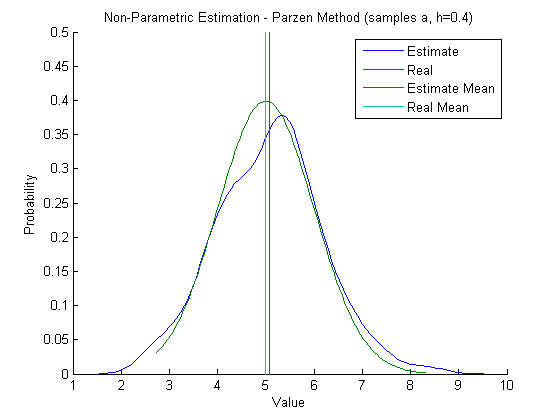


Figure : Parzen Estimation on Normal Data, h =0.4

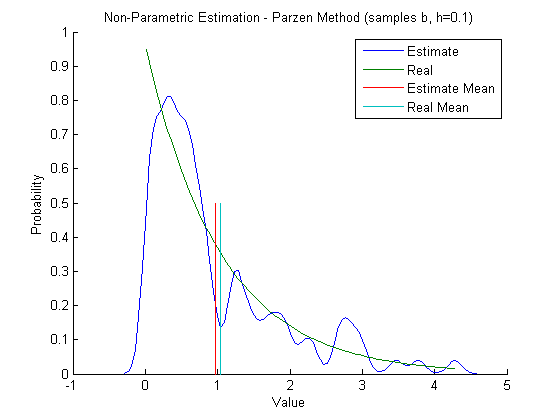


Figure : Parzen Estimation on Exponential Data, h =0.1

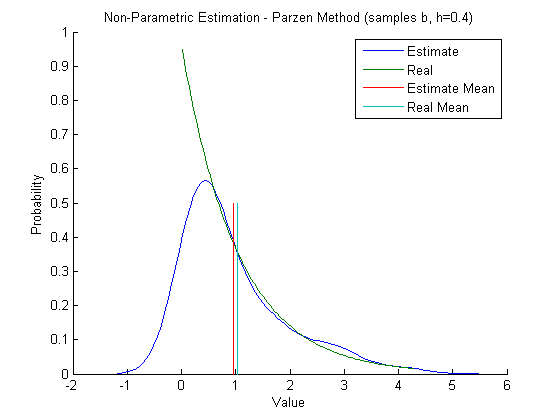


Figure : Parzen Estimation on Exponential Data, h =0.4

Observing the four graphs above it seems Parzen estimation is better suited for data set a, although at a greater value of h it can match to data set b decently at lower probability values.

In general however it is not possible always to use a parametric approach as the estimation assumption must be correct to the data set being analyzed which may not be possible all the time. It is better to use the parametric method when the distribution type of the dataset in question is known. The non-parametric method is preferred for any dataset that does not follow an exponential distribution as it can estimate much better for an entire range of value rather than just the lower probability ones.

# Model Estimation 2-D Case

