



Title of the Thesis

By

Name

Index No.: Index number

Dissertation

submitted in partial fulfillment of
the requirement for the degree of

Degree

in the

Department of Department name

Faculty of Applied Sciences

of the

University of Sri Jayewardenepura

Nugegoda

Sri Lanka.

Date

DECLARATION

I, Name, declare that this thesis titled, “Title of the Thesis” is the report of the research carried out under the supervision of Main Supervisor’s name as a partial fulfillment of the requirements of the course unit Course Code of the Degree at the University of Sri Jayewardenepura. It has not been submitted to any other institution or study program by me for any other purpose.

Signature:.....

Date:.....

SUPERVISOR'S RECOMMENDATION

I/We certify that this study was carried out by Name under my/our supervision.

Signature:.....

Date:.....

Main Supervisor's name,

Department of Department name,

University of Sri Jayewardenepura.

Signature:.....

Date:.....

2nd Supervisor's name

2nd Supervisor's Institute

Address Line 1

Address Line 2

Signature:.....

Date:.....

3rd Supervisor's name

3rd Supervisor's Institute

Address Line 1

Address Line 2

DEDICATION

Enter your dedication here

ACKNOWLEDGMENTS

Enter your acknowledgement here

ABSTRACT

Enter your abstract here. You may add keywords.

TABLE OF CONTENTS

DECLARATION	i
SUPERVISOR’S RECOMMENDATION	ii
DEDICATION	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	1
1 Introduction	2
1.1 Section name	2
1.1.1 subsection name	2
1.2 Background and Rationale	2
1.3 Introduction to the Mixture Distribution	2
1.4 Objectives of the Study	3

2	Literature Review	5
3	Methodology	6
4	Chapter name	7
4.1	Observations in General	8
5	Results	11
6	Conclusions and Discussion	12
	References	13
A	Tables	14
B	R codes	17
B.1	R implementation of the density(.) function	17

LIST OF TABLES

A.1	$N(160, 1), N(165, 4), N(170, 2)$ with proportions $1/4, 1/2, 1/4$	16
-----	--	-----------	----

LIST OF FIGURES

1.1	A mixture of three normal distributions with equal variances and with means 5, 10, and 15	3
4.1	Density curves of Caribbean (Barbados, Bermuda) based on h_AMISE bandwidth and epanechnikov kernel	8
4.2	A revealing map of the world's most and least ethnically diverse countries	9

CHAPTER 1

Introduction

1.1 Section name

1.1.1 subsection name

1.2 Background and Rationale

1.3 Introduction to the Mixture Distribution

Simply speaking, a mixture distribution is a probability distribution which consists 2 or more distributions where they mix in a way such that, area under the final density curve is equal to 1. More formally, we say that a distribution f given by equation 1.1 is a mixture of K component distributions f_1, f_2, \dots, f_K if

$$f(x) = \sum_{k=1}^K \lambda_k f_k(x) \quad (1.1)$$

with λ_k being the mixing weights $\lambda_k > 0$, $\sum_k \lambda_k = 1$ (Picard, 2007).

In probability and statistics, a mixture distribution is the probability distribution of a random variable that is derived from a collection of other random variables as follows: first, a random variable is selected by chance from the collection according to given probabilities of selection, and then the value of the selected random variable is realized. The underlying random variables may be random real numbers, or they may be random vectors (each having the same dimension), in which case the mixture distribution is a multivariate distribution (Frühwirth-Schnatter, 2006).

The distributions can be made up of different distributions (e.g. a normal distribution and a t-distribution) or they can be made up of the same distribution

with different parameters. For example, the following image (Figure 1.1) shows a mixture of three normal distributions (called a Gaussian Mixture Model), each with a different mean, but equal weight of $1/3$ and equal variance.

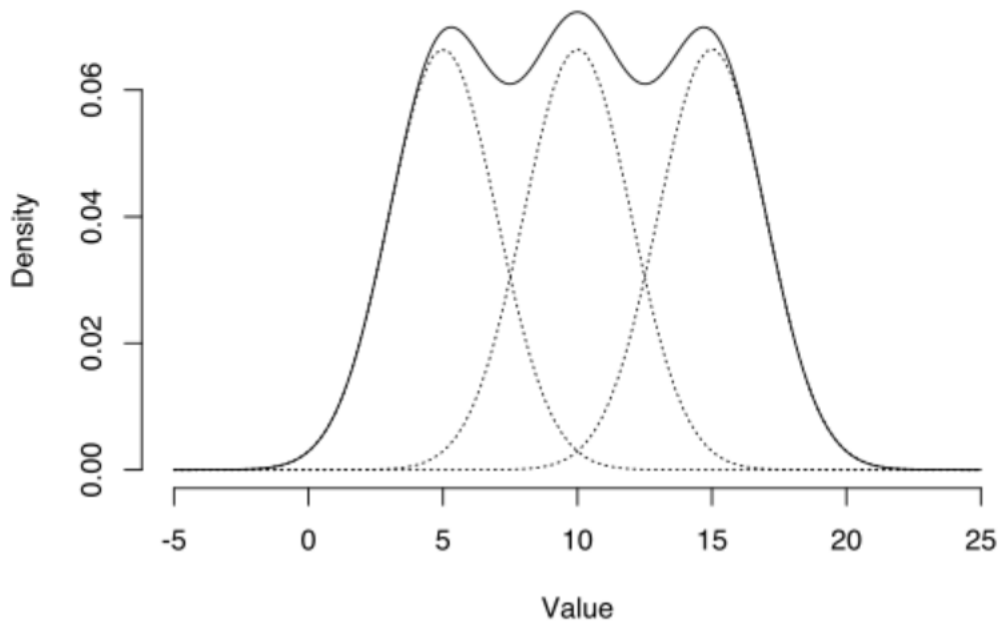


Figure 1.1: A mixture of three normal distributions with equal variances and with means 5, 10, and 15

1.4 Objectives of the Study

The objectives of the study are as follows:

1. Perform a comparison among several non-parametric density estimation methods utilizing literature.
2. Evaluate the performance of the Kernel Density Estimator for various existing bandwidth selection methods and kernel functions for various Gaussian mixtures.

3. Investigate on the possibility of improving the practical performance of the bandwidth selector to devise a more efficient bandwidth selection method.
4. Apply proposed estimation methods to estimate average adult human height distributions.
5. Analyze the distributions to identify modes of female and male adult height in various regions and how it has changed over the years (generations).

CHAPTER 2

Literature Review

CHAPTER 3

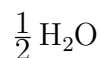
Methodology

CHAPTER 4

Chapter name

You may have several chapters depending on your study...

If you need to write chemical formula, then use the following format:



Adding sub figures could be done as follows:

Figure 4.1 reveals clear bi-modality throughout the first 4 generations. The latest cohort shows a slight tri-modality in the main 2 modes. There's a shift of the distributions to the right as years move ahead. Seemingly, it could be assumed that this population is homogeneous (with respect to heights of the ethnic groups living) and this bi-modality may be due to the gender for the first 4 cohorts.

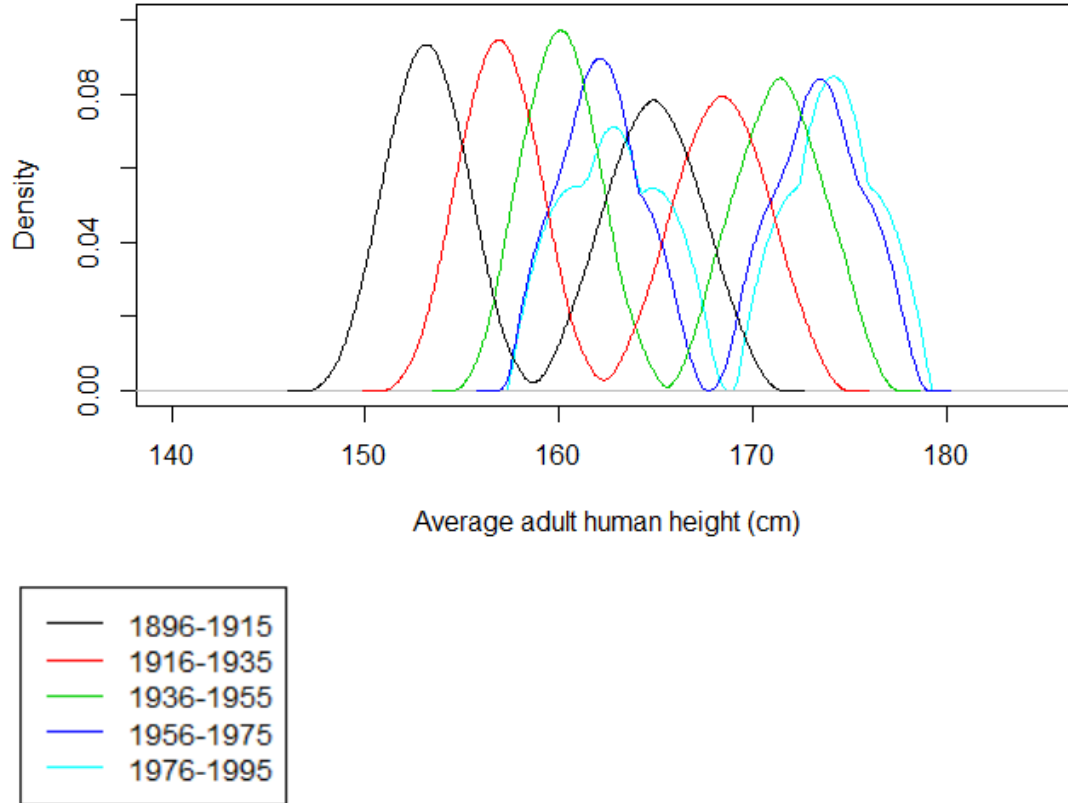


Figure 4.1: Density curves of Caribbean (Barbados, Bermuda) based on h_AMISE bandwidth and epanechnikov kernel

4.1 Observations in General

By plotting all 5 densities of cohort data in one plot for each region provides us with an important observation. That is, when the time goes forward, the density curves shift to the right. Which means, irrespective of the region, humans have grown taller with the time.

In most of the regions, it could clearly be seen a bimodal shape, which we can

give reasoning, that this happens due to gender discrimination. That is, a mixture of 2 gender populations. But in some cases, we can see more than 2 modes. That reveals, there are more than 2 mixtures of populations. Especially, regions which consist of countries with several ethnic groups show this feature.

Refer Figure 4.2 which provides evidence on how the regions of the world are ethnically diversified.

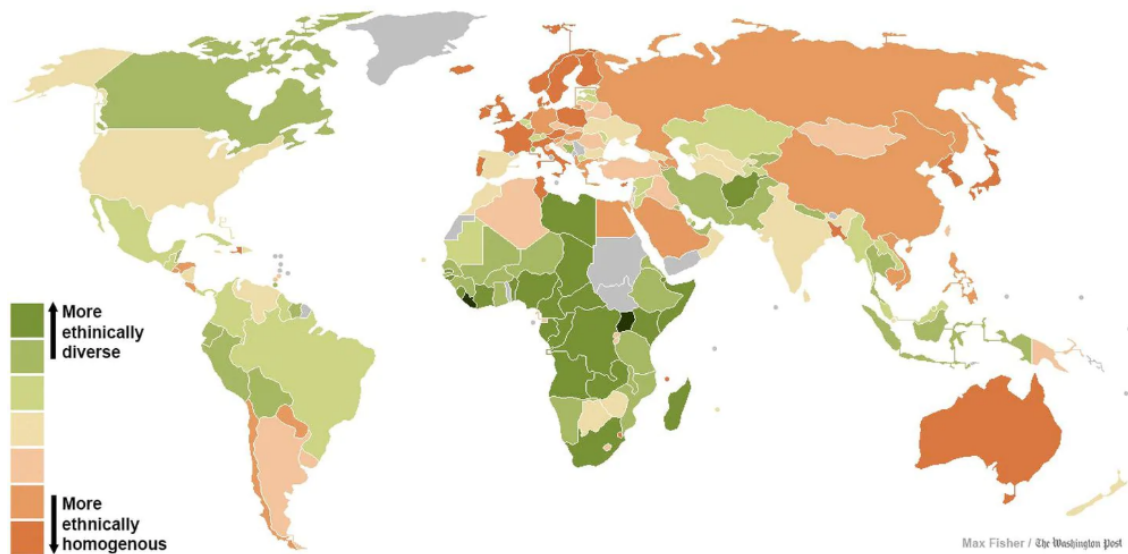


Figure 4.2: A revealing map of the world's most and least ethnically diverse countries
Source: Harvard Institute for Economic Research (2013)

The greener countries are more ethnically diverse and the orange countries more homogeneous. There are a few trends which could be seen right away: countries in Europe and Northeast Asia tend to be the most homogeneous, sub-Saharan African nations the most diverse. The Americas are generally somewhere in the middle. And richer countries appear more likely to be homogeneous. This provides sufficient

evidence to conclude that, mixtures with more than 2 modes are non-homogeneous regions with distinct ethnic groups living in.

CHAPTER 5

Results

CHAPTER 6

Conclusions and Discussion

References

- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Verlag New York.
- Picard, F. (2007). An introduction to mixture models. *Statistics for Systems Biology Group. Research Report*, (7).

APPENDIX A Tables

Sample Size	Bandwidth Selection Method	Kernel	Avg. MSE	Var. MSE	Avg. SSE % at Boundaries
20	Silverman's	Gaussian	0.0005196	3.37E-08	15.52559444
		Epanechnikov	0.0005775	2.75E-08	15.90946913
		Rectangular	0.0006603	2.72E-08	16.19555299
		Triangular	0.0005388	3.07E-08	15.6828454
		Cosine	0.0005493	2.96E-08	15.74082975
	Sheather and Jones	Gaussian	0.0005467	7.35E-08	12.7953963
		Epanechnikov	0.0005676	6.43E-08	13.45575486
		Rectangular	0.0006259	6.34E-08	13.71084743
		Triangular	0.0005512	6.80E-08	13.06122363
		Cosine	0.0005546	6.73E-08	13.21800687
	Unbiased Cross Validation	Gaussian	0.0006842	3.13E-07	12.71367795
		Epanechnikov	0.0006929	2.56E-07	13.45345975
		Rectangular	0.0007464	2.52E-07	13.89218775
		Triangular	0.0006802	2.77E-07	13.02377715
		Cosine	0.0006839	2.73E-07	13.15648133

Table A.1 – *Continued on next page*

50	Silverman's	Gaussian	0.0003413	1.02E-08	7.146318885
		Epanechnikov	0.0003896	8.64E-09	7.699787371
		Rectangular	0.0004446	8.53E-09	7.953545646
		Triangular	0.0003616	9.39E-09	7.542852767
		Cosine	0.0003674	9.18E-09	7.495687251
	Sheather and Jones	Gaussian	0.0003113	1.82E-08	4.615079558
		Epanechnikov	0.00033	1.82E-08	4.932247009
		Rectangular	0.0003635	1.88E-08	5.099644142
		Triangular	0.0003191	1.80E-08	4.806516274
		Cosine	0.0003204	1.81E-08	4.821287383
	Unbiased Cross Validation	Gaussian	0.0003634	4.99E-08	4.390166853
		Epanechnikov	0.000373	4.50E-08	4.687417995
		Rectangular	0.0004022	4.53E-08	4.872334205
		Triangular	0.0003659	4.64E-08	4.566824169
		Cosine	0.0003667	4.62E-08	4.580223994
100	Silverman's	Gaussian	0.0002401	3.60E-09	3.330362725
		Epanechnikov	0.0002809	3.12E-09	3.458360128
		Rectangular	0.0003219	3.09E-09	3.415155759
		Triangular	0.0002588	3.36E-09	3.510091712
		Cosine	0.0002627	3.29E-09	3.438681256
	Sheather and Jones	Gaussian	0.0001851	7.06E-09	1.655168036
		Epanechnikov	0.0001947	7.16E-09	1.63788748
		Rectangular	0.0002125	7.23E-09	1.636279701
		Triangular	0.0001897	7.07E-09	1.640550852
		Cosine	0.0001899	7.09E-09	1.647035341
	Unbiased Cross Validation	Gaussian	0.0002138	1.50E-08	1.541795705
		Epanechnikov	0.0002156	1.37E-08	1.558855769
		Rectangular	0.00023	1.35E-08	1.581101333
		Triangular	0.0002137	1.41E-08	1.546909474
		Cosine	0.0002136	1.40E-08	1.551008455

Table A.1 – *Continued on next page*

10000	Silverman's	Gaussian	1.47E-05	6.62E-12	0.000856781
		Epanechnikov	1.62E-05	6.87E-12	0.000760969
		Rectangular	1.77E-05	8.16E-12	0.00076494
		Triangular	1.57E-05	6.79E-12	0.000786132
		Cosine	1.57E-05	6.80E-12	0.000783879
	Sheather and Jones	Gaussian	3.76E-06	1.69E-12	0.006853138
		Epanechnikov	3.66E-06	1.70E-12	0.006278352
		Rectangular	6.22E-06	6.27E-12	0.003676873
		Triangular	3.66E-06	1.67E-12	0.006753898
		Cosine	3.67E-06	1.69E-12	0.006517576
	Unbiased Cross Validation	Gaussian	3.89E-06	1.78E-12	0.006582121
		Epanechnikov	3.76E-06	1.74E-12	0.006106225
		Rectangular	6.89E-06	1.24E-11	0.003523187
		Triangular	3.78E-06	1.76E-12	0.006467651
		Cosine	3.78E-06	1.75E-12	0.006319692

Table A.1: $N(160, 1)$, $N(165, 4)$, $N(170, 2)$ with proportions $1/4, 1/2, 1/4$

APPENDIX B

R codes

B.1 R implementation of the density(.) function

```
1 library(kedd)
2 dens <- function(x_mix){
3   K <- c()
4   d <- c()
5   x_min = min(x_mix)
6   x_max = max(x_mix)
7   h.deri = bw.nrd0(x_mix)
8   d_der12<-dkde(x_mix,c(seq(x_min-h.deri, x_max+h.deri,
9                               by=0.1)),h=h.deri,kernel="epanechnikov",
10                      deriv.order=2)$est.fx
11  d_der12_sq<-d_der12^2
12  tot_d_der12<-sum(d_der12_sq)*0.1
13  h_AMISE<-((3/5)/(1/25*length(x_mix)*tot_d_der12))^(1/5)
14  n = length(x_mix)
15  grid<-c(seq(x_min-h_AMISE, x_max+h_AMISE, by = 0.1))
16  for (i in 1:length(grid)) {
17    for (j in 1:n) {
18      u = (grid[i] - x_mix[j])/(h_AMISE)
19      if(abs(u) <= sqrt(5)){
20        K[j] = sqrt(1/5)*3/4*(1-1/5*u^2)}
21      else{K[j] = 0}}
22    d[i] = sum(K)/(h_AMISE*n)}
23  return(d)}
```