BAYESIAN MIXTURE MODELS FOR DENSITY ESTIMATION

A Thesis in
Statistics
by
Huei-Wen Teng

© 2009 Huei-Wen Teng

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2009
The thesis of Huei-Wen Teng was reviewed and approved* by the following:

John C. Liechty  
Associate Professor of Marketing and Statistics

Bruce G. Lindsay  
Professor of Statistics  
Statistics Department Head

Murali Haran  
Assistant Professor of Statistics

Zhibiao Zhao  
Assistant Professor of Statistics

*Signatures are on file in the Graduate School.
Abstract

Mixture models provide a convenient and flexible class of models for density estimation. They are typically used to model data generated from one of a number of different groups. This thesis studies two types of mixture models for density estimation from a Bayesian perspective. First, the Dirichlet process mixture (DPM) model is reviewed as it allows flexible nonparametric modeling. Second, the Markov chain Monte Carlo, Monte Carlo (MC3) method is proposed. The idea behind the MC3 method is to approximate an unknown density with a discrete distribution that has a finite number of support points. From a modeling perspective, MC3 represents a finite mixture model where the mixture components are equally weighted. An efficient slice sampling algorithm is provided to implement the MC3 method. Simulation results show that these two methods produce comparable predictive densities. Differences between the DPM and MC3 from modeling and computation aspects are discussed. We conclude with a discussion of applying the MC3 approach to inference for integral equations, which include the state-price density estimation in finance as a specific example.
Table of Contents

List of Figures vi
List of Tables vii
Acknowledgments viii

Chapter 1
Introduction 1
1.1 Mixing density estimation 1
1.1.1 EM-based algorithms 1
1.1.2 Deconvolutions and the kernel density estimation 2
1.1.3 Mixture models 3
1.2 The problem setting 3
1.3 Bayesian solutions to density estimation 4
1.4 Outline 4

Chapter 2
The Dirichlet process mixture model 5
2.1 Introduction to the Dirichlet Process 5
2.2 Introduction to Dirichlet process mixture (DPM) models 6
2.3 Gibbs sampling method of DPM 7

Chapter 3
The Markov chain Monte Carlo, Monte Carlo Approach 9
3.1 Introduction to the Markov chain Monte Carlo, Monte Carlo (MC3) method 9
3.2 An efficient algorithm for the MC3 method 10
3.2.1 Introduction to slice sampling 10
3.2.2 Gibbs sampling for the MC3 method using slice sampling . . 12

Chapter 4
Numerical results 15
  4.1 Study design .................................. 15
  4.2 Simulation results ............................ 17

Chapter 5
Conclusions and final remarks 26
  5.1 Conclusions ..................................... 26
  5.2 Final remarks .................................... 26

Appendix A
Dataset 1 29
  A.1 $X \sim N(0,9)$ ................................... 29

Appendix B
Dataset 2 31
  B.1 $X \sim 0.3N(2,2) + 0.7N(9,2)$ .................... 31

Bibliography 33
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Predictive Density and 80% Credible Region: DPM and Dataset 1</td>
<td>18</td>
</tr>
<tr>
<td>4.2</td>
<td>KL Distance: DPM and Dataset 1</td>
<td>19</td>
</tr>
<tr>
<td>4.3</td>
<td>Predictive Density and 80% Credible Region: MC3 and Dataset 1</td>
<td>20</td>
</tr>
<tr>
<td>4.4</td>
<td>KL Distance: MC3 and Dataset 1</td>
<td>21</td>
</tr>
<tr>
<td>4.5</td>
<td>Predictive Density and 80% Credible Region: DPM and Dataset 2</td>
<td>22</td>
</tr>
<tr>
<td>4.6</td>
<td>KL Distance: DPM and Dataset 2</td>
<td>23</td>
</tr>
<tr>
<td>4.7</td>
<td>Predictive Density and 80% Credible Region: MC3 and Dataset 2</td>
<td>24</td>
</tr>
<tr>
<td>4.8</td>
<td>KL Distance: MC3 and Dataset 2</td>
<td>25</td>
</tr>
<tr>
<td>A.1</td>
<td>Realized data and true density for Dataset 1</td>
<td>29</td>
</tr>
<tr>
<td>B.1</td>
<td>Realized data and true density for Dataset 2</td>
<td>31</td>
</tr>
</tbody>
</table>
List of Tables

4.1 Distributions of the Two Datasets .................................................. 16
4.2 Summary of the Datasets ............................................................... 16
4.3 Summary of Predictive Densities of Dataset 1 ................................. 17
4.4 Summary of Predictive Densities of Dataset 2 ................................. 18
A.1 Dataset 1 .................................................................................. 30
B.1 Dataset 2 .................................................................................. 32
Acknowledgments

I gratefully acknowledge my advisor Dr. John Liechty for his guidance supervision and crucial contribution. It is also my pleasure to convey my gratitude to my committee members, Dr. Jay Huang, Dr. Murali Haran, and Dr. Zhibiao Zhao, for their contributions to this thesis, their suggestions and kind discussions.

I will also thank the statistics department at Penn State University, where I received high level education, and found my interests in quantitative analysis. I have taken courses with Dr. Bruce Lindsay, Dr. Naomi Altman, Dr. Murali Haran, Dr. Runze Li, Dr. David Hunter, Dr. Steve Arnold, Dr. John Fricks, Dr. Bing Li, Dr. Debashi Goshi, Dr. Yu Zhang. Special thanks to Dr. Bruce Lindsay who was my temporary advisor for my first few years as a graduate student and helped me to get accustomed to graduate life. I also thank Dr. James Rosenberger, Dr. Donald Richards, Dr. William Harkness, Dr. Naomi Altman, and Dr. Runze Li, for their great kindness and support in my research and in my life.

Finally, the list of thanks would not be complete without a mention of my family, to whom I owe a great deal and more. This is for my family.
Dedication

This thesis is dedicated to my dearest family.
1.1 Mixing density estimation

Density estimation is important and has been widely studied for many years. Assume that $Y_1, \ldots, Y_n$ are identically and independently distributed with a mixture density

$$h(y|G) = \int f(y|x) dG(x). \quad (1.1)$$

The function $G$ is called the mixing (latent) distribution. Estimating the unknown mixing distribution $G$ when $f(y|x)$ is a known component density is an old problem. When the density of $X$ exists, Eq. (1.1) becomes

$$h(y|G) = \int f(y|x) g(x) dx, \quad (1.2)$$

where $g$ is called the mixing (or latent) density.

The non-parametric maximum likelihood estimate (NPMLE) for $g$ is discrete with at most $n$ mass points; see Laird (1978) and Lindsay (1983). However, when $g$ is known to be absolutely continuous, the NPMLE is degenerate and hence is improper. Existing methods to estimate the mixing density are reviewed as follows.

1.1.1 EM-based algorithms

Vardi and Lee (1993) reformulated the mixing density estimation as a linear inverse problem with positivity restrictions (a LININPOS problem for short). LININPOS
problems consider
\[ h(y) = \int f(x, y)g(x)dx, \quad (1.3) \]
where \( h, f \) and \( g \) are non-negative. Here, \( f \) is assumed to be known and \( g \) needs to be estimated. LINIPOS problems cover classical problems, such as algebraic systems of linear systems of linear equations, Fredholm’s integral equations of the first kind, mixture models, deconvolutions, and so on. The authors also proposed a continuous EM algorithm to find the mixing density \( g \). Other EM-Based algorithms include, for example, the smooth-by-roughening method by Laird and Louis (1991), the smoothed EM algorithm by Silverman et al. (1990), and the doubly smoothed EM algorithm by Szkutnik (2003).

1.1.2 Deconvolutions and the kernel density estimation

When \( f(y|x) \) is a location density \( f(y - x) \), Eq. (1.2) becomes
\[ h(y|G) = \int f(y - x)g(x)dx, \quad (1.4) \]
which is known as the convolution of \( f \) and \( g \). For the deconvolution problem, another equivalent formulation is
\[ Y = X + \varepsilon, \quad (1.5) \]
where the observation \( Y \) is the sum of a random variable of interest \( X \) and an error. The density for \( Y \) in Eq. (1.5) is obtained using the convolution formula where \( X \) has density \( g \) and \( \varepsilon \) has density \( f \), and hence equals Eq. (1.4). This is the reason why estimating the unknown density \( g \) is called deconvolution.

Kernel methods can be applied in the deconvolution case. Kernel density estimation is one of the most popular nonparametric density estimation. If the observations \( Y_i \) are observed without error, \( g \) can be estimated by
\[ \hat{g}(x) = \frac{1}{nw} \sum_{i=1}^{n} K \left( \frac{x - Y_i}{w} \right), \quad (1.6) \]
where \( K(x) \) is a kernel density function chosen by the researcher, and \( w \) is the
bandwidth (Silverman, 1981). The bandwidth decides the smoothness of the resulting curve and hence its selection is crucial. For automatic bandwidth selection, see Chiu (1992). While kernel density estimation is widely used, it doesn’t work in more than one or two dimensions, and in practice it requires a huge amount of data to cover the true density. Basic references about kernel density estimation include Stefanski and Carroll (1990), Zhang (1990), and Fan (1991).

1.1.3 Mixture models

Recall the mixture density in Eq. (1.1)

$$h(y|G) = \int f(y|x)dG(x).$$

If $G$ is modeled without parametric assumptions, the density $h(y|G)$ is a semiparametric mixture. If $G$ is degenerate, say, discrete with $k$ points of support, Eq. (1.1) is a $k$-component mixture model (Lindsay and Lesperance, 1995).

Mixture models have been applied widely since Pearson (1894), where a method of moments approach was proposed to estimate the parameters of a two-component mixture model. For a comprehensive list of applications using mixture models, please see Titterington et al. (1985) and Titterington (1997). For density estimation using mixture models, see Roeder (1990) and Priebe (1994).

Recently, mixtures models are more commonly calibrated using maximum likelihood or Bayesian methods. For a summary of the non-Bayesian approach, see Titterington et al. (1985) and McLachlan and Basford (1988). Bayesian methods have become popular because of the developments in methodology and computation power. For a summary of the Bayesian approach to mixture models, see Deibolt and Robert (1994), Robert (1994), Gelman et al. (1995), Robert (1996), Escobar and West (1995), Richardson and Green (1997), and Stephens (1997).

1.2 The problem setting

To describe the hierarchical structure of the model, we consider the following mixture model. Suppose the data, $Y_1, \ldots, Y_n$, are conditionally independent and
normally distributed.

\[ Y_i | X_i \sim N(X_i, \sigma^2) \quad i = 1, \ldots, n, \]
\[ X_i | G \sim G \quad i = 1, \ldots, n. \]  \hspace{1cm} (1.7)

We note that Model (1.7) is a special case of Eq. (1.1) when \( f(y|x) \) is a normal distribution with mean \( (y - x) \) and variance \( \sigma^2 \), and it is also a special case of the deconvolution problem in Eq. (1.5) when the error \( \varepsilon \) is normally distribution with mean zero and variance \( \sigma^2 \). The mean values \( X_i \)'s come from the prior distribution, \( G \). Estimation and inference in the general mixture model depends on \( G \).

1.3 Bayesian solutions to density estimation

Let \( D_n = \{y_1, \ldots, y_n\} \) be the observed data. And let \( p(\cdot|\cdot) \) denote the conditional density. For density estimation, the Bayesian approach focuses on (1) inference on the mixing distribution \( G \), and (2) computing the predictive distribution of the “next” observation. This thesis focuses on the latter case, where the Bayesian density estimation problem (or predicting problem) is solved by summarizing the unconditional predictive density, \( p(y_{n+1}|D_n) \). The predictive density is usually not applicable in closed form and needs to be approximated using numerical methods.

1.4 Outline

The remainder of this thesis is organized as follows. Chapter 2 reviews the Dirichlet process, and its extension, the Dirichlet process mixture model. For computation purposes, a Polya urn Gibbs sampling method proposed by Escobar (1994) is reviewed. Chapter 3 introduces a Markov chain Monte Carlo, Monte Carlo method, which represents a finite mixture model, where each component has equal weight. For this method, a Gibbs sampling method using slice sampling is proposed. Chapter 4 contains a simulation study. Conclusions and final remarks are given in Chapter 5.
Chapter 2

The Dirichlet process mixture model

2.1 Introduction to the Dirichlet Process

The Dirichlet process was proposed by Ferguson (1973) as an approach to make nonparametric inference. Let $DP(\alpha, G_0)$ denote a Dirichlet process with two parameters, a location parameter, $G_0$, which is a probability measure on the $x$ parameter space, and a dispersion parameter $\alpha > 0$. The parameter $G_0$ is the mean distribution for the Dirichlet process, and the parameter $\alpha$ is a measure of the strength in the belief that $G$ is equal to $G_0$ a priori.

Blackwell and MacQueen (1973) connected the Dirichlet process with a generalized Polya urn scheme, and showed that

$$X_1 \sim G_0,$$

$$X_i | X_1, \ldots, X_{i-1} \left\{ \begin{array}{l}
= X_j, \quad \text{with probability } \frac{1}{\alpha + i - 1}, \\
\sim G_0, \quad \text{with probability } \frac{\alpha}{\alpha + i - 1}.
\end{array} \right.$$ 

Define $\delta(X, \cdot)$ as a unit point measure by

$$\delta(X, B) = \left\{ \begin{array}{ll}
1, & \text{when } X \in B, \\
0, & \text{otherwise}.
\end{array} \right.$$
The above conditional distribution of $X_i|X_1,\ldots,X_{i-1}$ can be written as

$$X_i|X_1,\ldots,X_{i-1} \sim \frac{\alpha}{\alpha + i - 1} G_0 + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta(X_j, \cdot).$$

The closed form of the joint probability of $X_1,\ldots,X_n$ is

$$dF(X_1,\ldots,X_n) = \prod_{i=1}^{n} \frac{\alpha G_0(dX_i) + \sum_{j=1}^{i-1} \delta(X_j, dX_i)}{\alpha + i - 1}.$$

Denote $X = (X_1,\ldots,X_n)$ and $X_{(-i)} = (X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$. Blackwell and MacQueen also showed that $X_i$’s are exchangeable and exhibit a so-called clustering effect, i.e.,

$$X_i|X_{(-i)} \sim \frac{\alpha}{\alpha + n - 1} G_0 + \frac{1}{\alpha + n - 1} \sum_{j=1}^{n} \delta(X_j, \cdot).$$

The posterior distribution of a new $X_{n+1}$ given $X$ is

$$X_{n+1}|X \sim \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta(X_i, \cdot). \tag{2.1}$$

### 2.2 Introduction to Dirichlet process mixture (DPM) models

Model (1.7) leads to a Dirichlet process mixture model (DPM) when the mixing distribution $G$ is a Dirichlet process. Mixtures of Dirichlet process provide a natural setting for density estimation when data is modeled as a sample from a mixtures of normal distributions Antoinak (1974). The Dirichlet process assigns a distribution on distributions, and forms a popular class of Bayesian nonparametric methods. However, DPM is limited in its application in the past because it is analytically complicated. Efficient simulation methods can be found in Escobar and West (1995), MacEachern and Müller (1998), Neal (2000), and Idier (2008).
The complete model specifications for DPM in this thesis are given as follows.

\[
\begin{align*}
Y_i | X_i & \sim N(x_i, \sigma^2) \quad i = 1, \ldots, n, \\
X_i | G & \sim G \quad i = 1, \ldots, n, \\
G | \alpha, G_0 & \sim DP(\alpha, G_0), \\
G_0 & \sim N(\mu, \sigma^2).
\end{align*}
\] (2.2)

For simplicity, we fix parameters \(\alpha, \mu\) and \(\sigma^2\) as constants.

### 2.3 Gibbs sampling method of DPM

The symbols \(Y\) and \(X\) denote the vectors \((Y_1, \ldots, Y_n)\) and \((X_1, \ldots, X_n)\). To sample \(X\) from its posterior distribution under Model (2.2), Escobar (1994) derived the conditional distribution of \(X_i\) given \(X_{(-i)}\) and \(Y\) in the following form:

\[
dF(X_i | X_{(-i)}, Y) = \frac{\phi(Y_i - X_i) \alpha g_0(dx_i) + \sum_{j=1, \ldots, n, j \neq i} \phi(Y_i - X_j) \delta(X_j, dx_i)}{A(Y_i) + \sum_{j=1, \ldots, n, j \neq i} \phi(Y_i - X_j)},
\] (2.3)

where \(\phi\) is the standard normal density function, and \(A(Y)\) is defined as

\[
A(Y) = \alpha \phi(Y - x) g_0(dx).
\]

When \(X_i\) has conditional distribution in Eq. (2.3), it can be sampled using the following rule:

\[
X_i | X_{(-i)}, Y \left\{ \begin{array}{ll}
= X_j, & \text{with probability } \frac{\phi(Y_i - X_j) A(Y_i)}{A(Y_i) + \sum_{l=1, \ldots, n, l \neq i} \phi(Y_i - X_l)}, \\
\sim h(X_i | Y_i), & \text{with probability } \frac{A(Y_i)}{A(Y_i) + \sum_{l=1, \ldots, n, l \neq i} \phi(Y_i - X_l)},
\end{array} \right.
\]

where \(h\) is a density function from which to sample \(X_i\) and is defined as

\[
h(X_i | Y_i) = \frac{A_0}{A(Y_i)} \phi(Y_i - X_i) g_0(X_i)
\]
with \( g_0 \) being the probability density function corresponding to the probability measure \( G_0 \).

Let \( X_m^{(0)} \) denote the sample of \( X_m \) at the \( k \)-th iteration in the Markov chain Monte Carlo simulation. A Markov chain Monte Carlo algorithm for drawing samples of \( X \) that have the conditional distribution in Eq. (2.3) is as follows.

1. Initialization: Start with \( x^{(0)} \) arbitrary.

2. Step \( k \). For \( k = 1, \ldots, \)
   
   \begin{enumerate}
   \item[(a)] Sample \( x_1^{(k)} \) from \( X_1|X_2 = x_2^{(k-1)}, \ldots, X_n = x_n^{(k-1)}, Y \).
   \item[(b)] Sample \( x_2^{(k)} \) from \( X_2|X_1 = x_1^{(k)}, X_3 = x_3^{(k-1)}, \ldots, X_n = x_n^{(k-1)}, Y \).
   \item[(c)] \ldots
   \item[(d)] Sample \( x_n^{(k)} \) from \( X_N|X_1 = x_1^{(k)}, \ldots, X_{n-1} = x_{n-1}^{(k)}, Y \).
   \end{enumerate}

The conditional distribution of \( X_i|X_{(-1)}, Y \) has the following form,

\[
X_i|X_{(-1)}, Y \sim q_0 N \left( \frac{\mu + \sigma^2 Y_i}{\sigma^2 + 1}, \frac{\sigma^2}{\sigma^2 + 1} \right) + \sum_{j=1, \ldots, n, j \neq i} q_j \delta(X_j, \cdot),
\]

where

\[
q_0 = \frac{A(Y_i)}{A(Y_i) + \sum_{l \neq i} \phi(Y_i - X_l)},
\]

\[
q_j = \frac{\phi(Y_i - X_j)}{A(Y_i) + \sum_{l \neq i} \phi(Y_i - X_l)} \quad j = 1, \ldots, n, j \neq i,
\]

\[
A(Y) = \frac{\alpha}{\sqrt{2\pi(\sigma^2 + 1)}} e^{-\frac{(Y - \mu)^2}{2(\sigma^2 + 1)}}.
\]
Chapter 3

The Markov chain Monte Carlo, Monte Carlo Approach

3.1 Introduction to the Markov chain Monte Carlo, Monte Carlo (MC3) method

Model (1.7) leads to a finite mixture model when $G$ is a discrete distribution with a finite number of support points. Now, consider a finite mixture model with $M$ ($M$ maybe unknown but finite) components as follows.

\[
Y_i | X_i \sim N(X_i, \sigma_i^2) \quad i = 1, \ldots, n,
\]
\[
X_i | G \sim G \quad i = 1, \ldots, n,
\]

where

\[
G | p, \pi = p_1 \delta(\pi_1, \cdot) + \cdots + p_M \delta(\pi_M, \cdot),
\] (3.1)

with $p = (p_1, \ldots, p_m)$ being the mixture components which are constrained to be non-negative and sum up to unity; and $\pi = (\pi_1, \ldots, \pi_M)$ being the component specific parameters with $\pi_m$ being specific to component $m$. To complete the model specifications, we still need to place a prior distribution on $p$ and $\pi$. A common prior distribution is

\[
\pi_m | G_0 \sim G_0 \quad m = 1, \ldots, M,
\]
\[ p|\delta \sim D(\delta_1, \ldots, \delta), \]

where \( D(\delta_1, \ldots, \delta_M) \) denotes the Dirichlet distribution with parameters \((\delta_1, \ldots, \delta_M)\).

When \( p_m \) is fixed at \( 1/M \) for \( m = 1, \ldots, M \), Model (3.1) reduces to

\[
\begin{align*}
Y_i|X_i & \sim N(X_i, \sigma^2) \quad i = 1, \ldots, n, \\
X_i|G & \sim G(\cdot) \quad i = 1, \ldots, n, \\
G|\pi & = \frac{1}{M} \delta(\pi_1, \cdot) + \cdots + \frac{1}{M} \delta(\pi_M, \cdot), \\
\pi_m|G_0 & \sim G_0 \quad m = 1, \ldots, M.
\end{align*}
\]

The likelihood for Model (3.2) is

\[
L(y|\pi) = \prod_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \pi_m)^2}{2\sigma^2}}.
\]

As a result, the component specific parameters \( \pi \) have two features:

1. A priori \( \pi \) represent draws from \( G_0 \) in a Monte Carlo simulation and are used to calculate the likelihood \( L(y|\pi) \),

2. A posterior \( \pi \) are parameters which need to be retained and updated in the Markov chain Monte Carlo simulate.

As a result, we call the method based on the model specified in Model (3.2), a Markov chain Monte Carlo, Monte Carlo (MC3), method.

### 3.2 An efficient algorithm for the MC3 method

#### 3.2.1 Introduction to slice sampling

Suppose we want to sample \( x \) with density proportional to some function \( f(x) \). For simplicity, we assume here that \( x \) is univariate. We can do this by introducing an auxiliary variable, \( u \), and defining a joint distribution over \( x \) and \( u \) as follows.

\[
p(x, u) \propto 1_{\{0 < u < f(x)\}},
\]

where \( \pi \) denote the component-specific parameters.
where \(1_{\{A\}}\) is an indicator function with support \(A\). That is to say, \((x, u)\) is uniformly distributed over the region \(R = \{(x, y) : 0 < u < f(x)\}\). Specifically, the joint distribution \(p(x, u)\) is

\[
p(x, u) = \begin{cases} \frac{1}{\int f(x) dx} & \text{if } 0 < u < f(x), \\ 0, & \text{otherwise.} \end{cases}
\]

The marginal distribution of \(x\) equals

\[
p(x) = \int_0^{f(x)} \frac{1}{\int f(x) dx} du = \frac{f(x)}{\int f(x) dx} \propto f(x),
\]

as desired. Intuitively, we can sample from the joint density for \((x, u)\) using the Markov chain Monte Carlo algorithm and ignore \(u\) to get samples from the marginal density \(p(x)\). If we let \(x^{(k)}\) denote the sample of \(x\) at the \(k\)-th iteration in the Markov chain Monte Carlo simulation, the implementation of the slice sampler is summarized as follows.

1. Initialization: Start with a random \(x^{(0)}\).

2. For \(k = 1, \ldots, \)

   (a) Generate \(u^{(k)}\) uniformly over the interval \((0, f(x^{(k-1)}))\).

   (b) Generate \(x^{(k)}\) uniformly over

   \[
   S = \{x : 0 < u^{(k)} < f(x)\}. \tag{3.4}
   \]

We remark that the set \(S\) in Eq. (3.4) is referred to the “slice” generated by \(u\).

Edwards and Sokal (1988) generalize the above idea and propose an auxiliary variable method (also known as the Swendsen-Wang algorithm) for the Ising model. In their scheme, the density of \(x\) is proportional to a product of \(k\) nonnegative functions \(f_i(x)\). That is,

\[
p(x) \propto \prod_{i=1}^{k} f_i(x). \tag{3.5}
\]

They introduce \(k\) auxiliary variables, \(u_1, \ldots, u_k\) and define a joint distribution for \((x, u_1, \ldots, u_k)\) which is uniform over the region in which \(0 < u_i < f_i(x)\) for
\(i = 1, \ldots, k\). We can use the slice sampler to generate samples from, \((x, u_1, \ldots, u_n)\), given in Eq. (3.5) using the following steps:

1. Initialization: Start with a random \(x^{(0)}\).

2. For \(k = 1, \ldots,\)
   
   (a) Generate \(u_i^{(k)}\) uniformly over the interval \((0, f_i(x^{(k-1)}))\) for \(i = 1, \ldots, k\).
   
   (b) Generate \(x^{(k)}\) uniformly over the slice \(S\), which is the intersection of \(S_i\), with
   
   \[S_i = \{x : 0 < u_i^{(k)} < f_i(x)\}\] for \(i = 1, \ldots, k\).

Concurrently, Damien et al. (1999) view methods based on multiple auxiliary variables as a general approach to constructing Markov chain for Bayesian inference problems.

As a remark, it is usually safer to calculate \(g_i(x) = \log(f_i(x))\) instead of \(f_i(x)\) in order to avoid possible problems with floating-point underflow. When a log transformation is used, the auxiliary variable becomes \(z_i \sim g_i(x^0) - e_i\), where \(e_i\) is independently and identically distributed as an Exponential distribution with mean 1, written as \(\text{Exp}(1)\), for \(i = 1, \ldots, k\). Gibbs sampling is as follows.

1. Initialization: Start with a random \(x^{(0)}\).

2. For \(k = 1, \ldots,\)
   
   (a) Generate \(z_i^{(k)} \sim g_i(x^{(k-1)}) - e_i\), where \(e_i \sim \text{Exp}(1)\) for \(i = 1, \ldots, k\).
   
   (b) Generate \(x^{(k)}\) uniformly over the slice \(S\), which is the intersection of \(S_i\), with
   
   \[S_i = \{x : z_i^{(k)} < g_i(x)\}\] for \(i = 1, \ldots, k\).

### 3.2.2 Gibbs sampling for the MC3 method using slice sampling

We present the slice sampling for the MC3 approach as follows. Let \(\phi_{\sigma^2}\) denotes the density of a normal distribution with mean of zero and variance of \(\sigma^2\). An algorithm to generate the posterior distribution for \(\pi\) in Model (3.2), assuming \(\sigma^2\) is known, is given as follows.
1. Initialization: Start with a random \( \pi^{(0)} \).

2. For \( k = 1, \ldots \),

(a) For \( m = 1, \ldots, M \), generate \( \pi_m^{(k)} \) uniformly distributed over the slice

\[
S = \bigcap_{i=1}^{N+1} S_i,
\]

where \( S_i \)'s are given below in Eqs. (3.7) and (3.8). We show in the following that the slice \( S_i \) is an open interval. Recall that

\[
\pi_m | \cdots \sim \left( \prod_{i=1}^{N} \sum_{j=1}^{M} \phi_{\sigma^2} \left( Y_i - \pi_j \right) \right) p(\pi_m | \mu, \sigma^2),
\]

where \( | \cdots \) denote conditioning on the values of the remaining parameters and the data. Define

\[
g_i(\pi) = \log \left( \phi_{\sigma^2} \left( Y_i - \pi \right) + \sum_{j=1, \ldots, M, j \neq m} \phi_{\sigma^2} \left( Y_i - \pi_j^{(k-1)} \right) \right) \quad i = 1, \ldots, N,
\]

\[
g_{N+1}(\pi) = \log(p(\pi | \mu, \sigma^2)).
\]

Generate \( z_i \sim g_i(\pi_m^{(k-1)}) - e_i \) where \( e_i \sim \text{Exp}(1) \) for \( i = 1, \ldots, N + 1 \). Define

\[
\lambda_{i-m} = \sum_{j=1, \ldots, M, j \neq m} \phi_{\sigma^2} \left( Y_i - \pi_j^{(k-1)} \right).
\]

Simple algebra gives

\[
S_i = \{ \pi : g_i(\pi) > z_i \} = \{ \pi : \log(\phi_{\sigma^2} \left( Y_i - \pi \right) + \lambda_{i-m}) > z_i \} = \{ \pi : \phi_{\sigma^2} \left( Y_i - \pi \right) > \exp^{z_i} \lambda_{i-m} \}.
\]

By the definition of \( z_i \),

\[
\exp^{z_i} \lambda_{i-m} = (\phi_{\sigma^2} \left( Y_i - \pi_m^{(k-1)} \right) + \lambda_{i-m}) \exp^{-e_i} \lambda_{i-m}
\]
can be less than or equal to zero when $e_i$ is large. When this is the case, $S_i = (-\infty, \infty)$. Otherwise, note that

$$\exp z_i - \lambda_{i-m} \leq (\phi_{\sigma^2} (Y_i - \pi^{(k-1)}_m) + \lambda_{i-m}) - \lambda_{i-m} \leq 1/\sqrt{2\pi \sigma^2}.$$ 

On the other hand, it is clear that $\phi_{\sigma^2} (Y_i - \pi) \leq 1/\sqrt{2\pi \sigma^2}$. As a result, \( \{ \pi : \phi_{\sigma^2} (Y_i - \pi) > \exp z_i - \lambda_{i-m} \} \) is a finite open interval and equals $(Y_i - \Delta_i, Y_i + \Delta_i)$ where

$$\Delta_i = \sqrt{-2\sigma^2 \log \left( \sqrt{2\pi \sigma^2 (\exp z_i - \lambda_{i-m})} \right)}.$$ \hspace{1cm} (3.6)

To sum up, we have $S_i$ in the following form,

$$S_i = (Y_i - \Delta_i, Y_i + \Delta_i)1_{\{\exp z_i - \lambda_{i-m} > 0\}}$$

$$+ (-\infty, \infty)1_{\{\exp z_i - \lambda_{i-m} \leq 0\}},$$ \hspace{1cm} (3.7)

for $i = 1, \ldots, N$, where $\Delta_i$ is defined in Eq. (3.6). Similarly, standard algebra gives

$$S_{N+1} = \{ x : g_{N+1}(\pi) > z_{N+1} \} \hspace{1cm} (3.8)$$

$$= (\mu - \Delta_{N+1}, \mu + \Delta_{N+1})$$

where

$$\Delta_{N+1} = \sqrt{-2\sigma^2 \log \left( \sqrt{2\pi \sigma^2 \exp z_{N+1}} \right)}.$$ \hspace{1cm} (3.9)
Chapter 4

Numerical results

We conduct a Monte Carlo study in order to compare the predictive densities generated by MC3 with the predictive densities generated by DPM. In the preceding sections, parameters, such as $\alpha$ and $G_0$ of DPM and $M$, $\mu$, and $\sigma^2$ of MC3 are assumed to be fixed. However, we remark that a prior distribution can be placed on these values allowing inference to come from the observed data, as it can be challenging in practice to select appropriate values for these parameters. The overall study design is described in Section 4.1. The results of the simulation study are contained in Section 4.2.

### 4.1 Study design

We consider two datasets, each with 100 observations, where $X_i$ is generated from the mixing distribution, and $Y_i$ is the sum of $X_i$ and a standard normal error. For the first dataset, $X_i$ comes from a normal distribution, and for the second dataset, $X_i$ comes from a mixture of normals with two components and unequal weights. Table 4.1 summarizes the distributions used to generate the data. In Appendixes A and B, Figures A.1 and B.1 graphically depict the data points and true densities, and Tables A.1 and B.1 report these actual values of the generated data.

To investigate the impact of parameter sensitivity, we use a range of values of $\alpha$ for DPM and a range of values of $M$ for MC3. For DPM, we take $\alpha$ either 0.01, 1, or 100. For MC3, we take $M$ either 4, 16, or 64. Using an Empirical Bayes method, we fix $G_0$ in DPM and $G_0$ in MC3 as a normal distribution with mean
Table 4.1. Distributions of the Two Datasets

<table>
<thead>
<tr>
<th></th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$N(0, 9)$</td>
<td>$0.3N(2, 2) + 0.7N(9, 2)$</td>
</tr>
<tr>
<td>$Y$</td>
<td>$N(10, 10)$</td>
<td>$0.3N(2, 3) + 0.7N(9, 3)$</td>
</tr>
<tr>
<td>Mean of $Y$</td>
<td>10</td>
<td>6.9</td>
</tr>
<tr>
<td>Variance of $Y$</td>
<td>10</td>
<td>13.29</td>
</tr>
</tbody>
</table>

Table 4.2. Summary of the Datasets

<table>
<thead>
<tr>
<th></th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>9.81</td>
<td>7.46</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>9.34</td>
<td>13.36</td>
</tr>
</tbody>
</table>

and variance equal to the sample mean and sample variance of the dataset. Table 4.2 summarizes the sample means and the sample variances of the two synthetic datasets.

In the Markov chain Monte Carlo simulation, we discard the first 1000 samples, and used the next 1000 samples for inference. To calculate the Kullback-Leibler (KL) distance between the true distribution of $Y$, $p$, and the predictive distribution at the $k$-th iteration, $q^{(k)}$, we first select a proper open interval $(l, u)$ which dominates all the observations. The interval $(l, u)$ is then discretized with points $x_0 < x_1 < \cdots < x_I$ and $x_0 = l$ and $x_I = u$ and the KL distance is approximated by

$$ KL(p, q^{(k)}) = \sum_{i=1}^{I} \log(p(x_i)) \frac{q^{(k)}(x_i)}{p(x_i)} $$

The Kullback-Leibler distance can be used to check the convergence of Markov chain Monte Carlo simulation and to evaluate how well the predictive densities fit the true densities. To visualize the predictive densities, we depict the mean and the 80% credible region of the predictive densities that result from the Markov chain Monte Carlo inference. Finally, posterior summaries of KL distances, posterior means and variances, are reported as a measure of model fit.
Table 4.3. Summary of Predictive Densities of Dataset 1

<table>
<thead>
<tr>
<th></th>
<th>DPM</th>
<th>MC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.14</td>
<td>0.32</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>100</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>Mean KL distance</td>
<td>0.14</td>
<td>0.32</td>
</tr>
<tr>
<td>Mean Predictive Dist.</td>
<td>9.72</td>
<td>10.35</td>
</tr>
<tr>
<td>Variance Predictive Dist.</td>
<td>8.75</td>
<td>12.83</td>
</tr>
<tr>
<td></td>
<td>9.80</td>
<td>9.88</td>
</tr>
<tr>
<td></td>
<td>9.86</td>
<td>9.33</td>
</tr>
<tr>
<td></td>
<td>10.35</td>
<td>9.84</td>
</tr>
<tr>
<td></td>
<td>9.88</td>
<td>9.54</td>
</tr>
</tbody>
</table>

4.2 Simulation results

Figures 4.1 to 4.8 show the predictive densities with 80% credible regions and KL distances obtained by DPM and MC3 using these two datasets. Plots of the KL distances show that samplers in Markov chain Monte Carlo simulation mix well. Hence, inference based on these iterations is meaningful. Tables 4.3 and 4.4 summarize the mean of the KL distances, which are calculated using the predictive densities and the true densities, and the means and variances of the predictive densities. These two tables show that DPM and MC3 produce comparable density estimations with regards to KL distances.

From a modeling perspective, the DPM and MC3 are mixture models but with different mixing distributions. Specifically, DPM uses a Dirichlet process as a mixing distribution, whereas MC3 uses a discrete distribution with a finite number of support points as a mixing distribution. A major difference between these two methods is in making inferences on the number of clusters in the dataset. It is known that the number of clusters in DPM depends heavily on α. When α is fixed, the DPM automatically forms a posterior distribution for the number of clusters. Hence the number of clusters is random for DPM. On the other hand, if M is fixed under the MC3 method, the number of clusters in the dataset is fixed.

From a computation perspective, DPM takes about the same time to run for different values of α. However, the computation time in MC3 grows linearly in M.
Table 4.4. Summary of Predictive Densities of Dataset 2

<table>
<thead>
<tr>
<th></th>
<th>DPM</th>
<th></th>
<th>MC3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>0.01</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Mean KL distance</td>
<td>0.10</td>
<td>0.08</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>Mean Predictive Dist.</td>
<td>7.46</td>
<td>7.46</td>
<td>7.18</td>
<td>7.25</td>
</tr>
<tr>
<td>Variance Predictive Dist.</td>
<td>13.27</td>
<td>13.24</td>
<td>23.73</td>
<td>14.23</td>
</tr>
</tbody>
</table>

Figure 4.1. Predictive Density and 80% Credible Region: DPM and Dataset 1
The true density is in bold line, the mean predictive is in bold dotted-dash line, and the 80% credible region is between the other two dotted-dash lines.
Figure 4.2. KL Distance: DPM and Dataset 1
Figure 4.3. Predictive Density and 80% Credible Region: MC3 and Dataset 1
The true density is in bold line, the mean predictive is in bold dotted line, and the 80% credible region is between the other two dotted lines.
Figure 4.4. KL Distance: MC3 and Dataset 1
Figure 4.5. Predictive Density and 80% Credible Region: DPM and Dataset 2
The true density is in bold line, the mean predictive is in bold dotted-dash line, and the 80% credible region is between the other two dotted-dash lines.
Figure 4.6. KL Distance: DPM and Dataset 2
Figure 4.7. Predictive Density and 80% Credible Region: MC3 and Dataset 2
The true density is in bold line, the mean predictive is in bold dotted line, and the 80% credible region is between the other two dotted lines.
Figure 4.8. KL Distance: MC3 and Dataset 2
Conclusions and final remarks

5.1 Conclusions

This thesis compares the density estimation using Bayesian mixtures models. The Dirichlet Process mixture (DPM) model uses a Dirichlet process as a mixing distribution, whereas the Markov chain Monte Carlo, Monte Carlo (MC3) method uses a finite discrete distribution where each mixing component is given equal weight. This thesis provides an efficient and easy-to-implement slice sampling scheme for the MC3 method. A simulation study shows that these two methods produce predictive densities of similar quality with regards to Kullback-Leibler distances to the true densities. Differences between the DPM and MC3 from computation and modeling aspects are discussed.

5.2 Final remarks

As deconvolution problems fall into the category of the LININPOS problems mentioned in Subsection 1.1.1, it is straightforward to explore the applicability of the MC3 method to a general inverse problem. Indeed, the MC3 method can be used to make inferences for a more general inverse problem, particularly, integral equations observed with errors. For an integral equation, we consider that the observation $Y$ comes from

$$Y_i = \int f_i(x)g(x)dx + \varepsilon_i,$$

(5.1)
for \( i = 1, \ldots, n \), where \( f_i(x) \) is a known function, \( g \) is the density of the random variable of interest, \( X \), and \( \varepsilon \) is the error. We assume that the functional form of the error distribution is known. We note that the error term does not have to be additive in Eq. (5.1). Because \( g \) is a density, Eq. (5.1) equals
\[
Y_i = E[f_i(X)] + \varepsilon_i, \quad (5.2)
\]
for \( i = 1, \ldots, n \), where \( E[\cdot] \) is the expectation operator over \( X \) with respect to density \( g(x) \). Hence, Eq. (5.2) clarifies that the observed data is the expectation of a function of \( X \) that is contaminated with error. The question of interest is to make inferences for the unknown density \( g \).

To motivate practical interest in this problem, consider an example from finance, inference of the state-price density as discussed in Ait-Sahalia and Lo (1998). Let \((u)^+\) be the positive function which gives \( u \) for positive \( u \) and 0 otherwise. Consider \( N_1 \) put options with strike prices \( c_{1j} \) and \( N_2 \) call options with strike prices \( c_{2j} \). The payoff function of a European option is
\[
((-1)^i(X - c_{ij}))^+
\]
for \( i = 1, 2, j = 1, \ldots, N_i \). The theoretical option price for an options is the discounted expectation of the payoff function under the risk-neutral probability, i.e.,
\[
e^{-rT}E_Q[((-1)^i(X - c_{ij}))^+]
\]
for \( i = 1, 2, j = 1, \ldots, N_i \) where \( E_Q \) is the expectation over \( X \) with unknown risk-neutral density. Since options prices are non-negative in an arbitrage-free market, we assume the observed option prices follow
\[
Y_{ijk} = e^{-rT}E_Q[((-1)^i(X - c_{ij}))^+]e^{\varepsilon_{ijk}}
\]
for \( i = 1, 2, j = 1, \ldots, N_i, k = 1, \ldots, N_{ij} \), where \( \varepsilon_{ijk} \) is independently and identically normally distributed with mean of zero and variance of \( \sigma^2_\varepsilon \).

The MC3 method can be applied to the state-price density estimation. Suppose the unknown risk-neutral density can be approximated by a discrete distribution with a finite number of support points, \( \pi = (\pi_1, \ldots, \pi_M) \). Further, we assume that
\( \pi_m \) comes from \( G_0 \) for \( m = 1, \ldots, M \). Then the theoretic option price is

\[
G_{ij} = G_{ij}(\pi) = e^{-rT} \sum_{m=1}^{M} \frac{1}{M} \left((-1)^i (\pi_m - c_{ij})\right) ^+
\]

for \( i = 1, 2, j = 1, \ldots, N_i \). The likelihood is

\[
L(y|\pi) = \prod_{i=1}^{2} \prod_{j=1}^{N_i} \prod_{k=1}^{N_{ij}} \frac{1}{\sqrt{2\pi \sigma_k^2}} e^{-\frac{(\log Y_{ijk} - \log G_{ij})^2}{2\sigma_k^2 v_{ijk}}},
\]

where \( v_{ijk} \) is the volume of the corresponding traded options and is taken as known. The model specification is complete when a prior distribution is placed on the unknown parameters. Similarly, Markov chain Monte Carlo simulation can be used for inferences. Hence, the MC3 approach to the state-price density estimation is straightforward. As the focus the this thesis remains on density estimation, we leave the estimation of state-price density as a future exercise.
Appendix A

Dataset 1

A.1 $X \sim N(0, 9)$

Figure A.1. Realized data and true density for Dataset 1
### Table A.1. Dataset 1

<table>
<thead>
<tr>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.3433</td>
<td>26</td>
<td>8.3444</td>
<td>51</td>
<td>12.0481</td>
<td>76</td>
<td>10.4702</td>
</tr>
<tr>
<td>2</td>
<td>14.9838</td>
<td>27</td>
<td>12.0928</td>
<td>52</td>
<td>10.5597</td>
<td>77</td>
<td>7.9741</td>
</tr>
<tr>
<td>3</td>
<td>15.8673</td>
<td>28</td>
<td>13.2710</td>
<td>53</td>
<td>9.8928</td>
<td>78</td>
<td>7.5091</td>
</tr>
<tr>
<td>4</td>
<td>11.9615</td>
<td>29</td>
<td>11.6124</td>
<td>54</td>
<td>10.2274</td>
<td>79</td>
<td>8.1791</td>
</tr>
<tr>
<td>5</td>
<td>5.9286</td>
<td>30</td>
<td>18.1012</td>
<td>55</td>
<td>8.5166</td>
<td>80</td>
<td>12.4141</td>
</tr>
<tr>
<td>6</td>
<td>11.0675</td>
<td>31</td>
<td>9.6542</td>
<td>56</td>
<td>15.7018</td>
<td>81</td>
<td>10.0800</td>
</tr>
<tr>
<td>7</td>
<td>8.2359</td>
<td>32</td>
<td>11.2428</td>
<td>57</td>
<td>9.3356</td>
<td>82</td>
<td>4.5003</td>
</tr>
<tr>
<td>8</td>
<td>6.9223</td>
<td>33</td>
<td>4.0605</td>
<td>58</td>
<td>5.5155</td>
<td>83</td>
<td>15.6152</td>
</tr>
<tr>
<td>9</td>
<td>12.0685</td>
<td>34</td>
<td>8.1549</td>
<td>59</td>
<td>9.1225</td>
<td>84</td>
<td>12.0178</td>
</tr>
<tr>
<td>10</td>
<td>11.4166</td>
<td>35</td>
<td>4.3671</td>
<td>60</td>
<td>8.2726</td>
<td>85</td>
<td>5.2902</td>
</tr>
<tr>
<td>12</td>
<td>8.9272</td>
<td>37</td>
<td>6.1017</td>
<td>62</td>
<td>8.8820</td>
<td>87</td>
<td>10.4861</td>
</tr>
<tr>
<td>13</td>
<td>7.2065</td>
<td>38</td>
<td>7.9612</td>
<td>63</td>
<td>7.5785</td>
<td>88</td>
<td>7.6786</td>
</tr>
<tr>
<td>14</td>
<td>4.3820</td>
<td>39</td>
<td>7.2175</td>
<td>64</td>
<td>4.9457</td>
<td>89</td>
<td>9.6691</td>
</tr>
<tr>
<td>15</td>
<td>5.3529</td>
<td>40</td>
<td>9.6303</td>
<td>65</td>
<td>8.5341</td>
<td>90</td>
<td>10.9779</td>
</tr>
<tr>
<td>16</td>
<td>6.5877</td>
<td>41</td>
<td>12.7805</td>
<td>66</td>
<td>5.3078</td>
<td>91</td>
<td>13.1074</td>
</tr>
<tr>
<td>17</td>
<td>8.7863</td>
<td>42</td>
<td>14.4753</td>
<td>67</td>
<td>9.5484</td>
<td>92</td>
<td>5.8433</td>
</tr>
<tr>
<td>18</td>
<td>10.7262</td>
<td>43</td>
<td>11.2783</td>
<td>68</td>
<td>12.7685</td>
<td>93</td>
<td>6.8614</td>
</tr>
<tr>
<td>19</td>
<td>10.5318</td>
<td>44</td>
<td>8.0584</td>
<td>69</td>
<td>17.0081</td>
<td>94</td>
<td>7.9401</td>
</tr>
<tr>
<td>20</td>
<td>6.2540</td>
<td>45</td>
<td>10.6054</td>
<td>70</td>
<td>11.3547</td>
<td>95</td>
<td>10.3491</td>
</tr>
<tr>
<td>21</td>
<td>12.0355</td>
<td>46</td>
<td>7.4072</td>
<td>71</td>
<td>11.1715</td>
<td>96</td>
<td>14.1047</td>
</tr>
<tr>
<td>22</td>
<td>11.3934</td>
<td>47</td>
<td>7.1715</td>
<td>72</td>
<td>13.1560</td>
<td>97</td>
<td>15.7282</td>
</tr>
<tr>
<td>23</td>
<td>14.3447</td>
<td>48</td>
<td>5.7502</td>
<td>73</td>
<td>11.4472</td>
<td>98</td>
<td>12.3266</td>
</tr>
<tr>
<td>24</td>
<td>8.3000</td>
<td>49</td>
<td>4.3729</td>
<td>74</td>
<td>13.7616</td>
<td>99</td>
<td>12.3067</td>
</tr>
</tbody>
</table>
Appendix B

Dataset 2

B.1 \( X \sim 0.3N(2, 2) + 0.7N(9, 2) \)

![Graph of real data and true density for Dataset 2]

Figure B.1. Realized data and true density for Dataset 2
### Table B.1. Dataset 2

<table>
<thead>
<tr>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.7891</td>
<td>26</td>
<td>5.3302</td>
<td>51</td>
<td>8.3447</td>
<td>76</td>
<td>9.6426</td>
</tr>
<tr>
<td>2</td>
<td>5.9597</td>
<td>27</td>
<td>11.8859</td>
<td>52</td>
<td>3.4907</td>
<td>77</td>
<td>9.7661</td>
</tr>
<tr>
<td>3</td>
<td>3.4194</td>
<td>28</td>
<td>8.7347</td>
<td>53</td>
<td>6.9040</td>
<td>78</td>
<td>7.6243</td>
</tr>
<tr>
<td>4</td>
<td>2.6993</td>
<td>29</td>
<td>12.2416</td>
<td>54</td>
<td>11.7074</td>
<td>79</td>
<td>0.5164</td>
</tr>
<tr>
<td>5</td>
<td>2.7279</td>
<td>30</td>
<td>-1.0422</td>
<td>55</td>
<td>10.0488</td>
<td>80</td>
<td>8.3909</td>
</tr>
<tr>
<td>6</td>
<td>9.7637</td>
<td>31</td>
<td>9.1227</td>
<td>56</td>
<td>2.0733</td>
<td>81</td>
<td>2.8387</td>
</tr>
<tr>
<td>7</td>
<td>5.4990</td>
<td>32</td>
<td>12.3745</td>
<td>57</td>
<td>9.2198</td>
<td>82</td>
<td>7.1227</td>
</tr>
<tr>
<td>8</td>
<td>7.9771</td>
<td>33</td>
<td>7.1177</td>
<td>58</td>
<td>9.5950</td>
<td>83</td>
<td>10.3181</td>
</tr>
<tr>
<td>9</td>
<td>11.0494</td>
<td>34</td>
<td>6.6457</td>
<td>59</td>
<td>10.1432</td>
<td>84</td>
<td>13.1308</td>
</tr>
<tr>
<td>10</td>
<td>6.3865</td>
<td>35</td>
<td>9.0485</td>
<td>60</td>
<td>4.1419</td>
<td>85</td>
<td>7.5034</td>
</tr>
<tr>
<td>11</td>
<td>11.3244</td>
<td>36</td>
<td>1.6474</td>
<td>61</td>
<td>5.8314</td>
<td>86</td>
<td>-0.1853</td>
</tr>
<tr>
<td>12</td>
<td>9.4605</td>
<td>37</td>
<td>5.8703</td>
<td>62</td>
<td>5.3716</td>
<td>87</td>
<td>12.2060</td>
</tr>
<tr>
<td>13</td>
<td>10.6105</td>
<td>38</td>
<td>7.9619</td>
<td>63</td>
<td>6.5172</td>
<td>88</td>
<td>11.5712</td>
</tr>
<tr>
<td>14</td>
<td>-0.4655</td>
<td>39</td>
<td>7.7479</td>
<td>64</td>
<td>9.8875</td>
<td>89</td>
<td>4.1013</td>
</tr>
<tr>
<td>15</td>
<td>8.5213</td>
<td>40</td>
<td>9.1906</td>
<td>65</td>
<td>9.0912</td>
<td>90</td>
<td>10.8265</td>
</tr>
<tr>
<td>16</td>
<td>10.2735</td>
<td>41</td>
<td>7.8796</td>
<td>66</td>
<td>6.8698</td>
<td>91</td>
<td>0.3323</td>
</tr>
<tr>
<td>17</td>
<td>6.8762</td>
<td>42</td>
<td>1.0232</td>
<td>67</td>
<td>2.0572</td>
<td>92</td>
<td>10.6457</td>
</tr>
<tr>
<td>18</td>
<td>3.0828</td>
<td>43</td>
<td>-0.5073</td>
<td>68</td>
<td>6.9922</td>
<td>93</td>
<td>10.1687</td>
</tr>
<tr>
<td>19</td>
<td>11.0269</td>
<td>44</td>
<td>8.9705</td>
<td>69</td>
<td>8.9736</td>
<td>94</td>
<td>14.3108</td>
</tr>
<tr>
<td>20</td>
<td>7.0158</td>
<td>45</td>
<td>10.6022</td>
<td>70</td>
<td>6.6505</td>
<td>95</td>
<td>7.1780</td>
</tr>
<tr>
<td>21</td>
<td>9.9660</td>
<td>46</td>
<td>4.4033</td>
<td>71</td>
<td>7.9168</td>
<td>96</td>
<td>9.5153</td>
</tr>
<tr>
<td>22</td>
<td>13.6117</td>
<td>47</td>
<td>2.8705</td>
<td>72</td>
<td>1.2497</td>
<td>97</td>
<td>7.9280</td>
</tr>
<tr>
<td>23</td>
<td>10.3399</td>
<td>48</td>
<td>1.2338</td>
<td>73</td>
<td>11.6824</td>
<td>98</td>
<td>9.7304</td>
</tr>
<tr>
<td>24</td>
<td>3.4093</td>
<td>49</td>
<td>11.0874</td>
<td>74</td>
<td>10.6406</td>
<td>99</td>
<td>6.5862</td>
</tr>
<tr>
<td>25</td>
<td>11.1198</td>
<td>50</td>
<td>7.0489</td>
<td>75</td>
<td>7.9346</td>
<td>100</td>
<td>11.3565</td>
</tr>
</tbody>
</table>
Bibliography


