Gaussian Mixture Optimization and a Revised Black-Litterman Model

Itiel Cohen - CSI-CUNY
Benard Neuhaus - MU
David Suarez - UNCG
Emily Zucker- NCSU

Faculty Mentor: Dr. Tao Pang
Graduate Student Mentor: Cagatay Karan

Contents

Abstract

Gaussian mixture distributions are often used to model the distribution of data when a single normal distribution will not provide a suitable fit. A normal distribution only matches the first and second moments of the data, but a Gaussian mixture can fit higher moments of data as well. We consider several existing methods for optimizing a Gaussian mixture, both in the univariate and multivariate case and develop a computationally efficient method for univariate quantile regression. Also, we present a revised version of the Black-Litterman model, a well known financial model used to combine the views of an investor with a neutral starting point. The revised model utilizes multivariate Gaussian mixture distributions rather than multivariate normal distributions, and thus allows for the preservation of the higher comoments of asset returns.

1 Introduction

Several different methods can be used to optimize the parameters of a univariate Gaussian mixture so that it fits a given data set. We will look at three different methods of fitting a univariate Gaussian mixture to data. Two of the methods are previously described in other works- the Method of Moments (Pearson 1894) and the Maximum Likelihood Method (Behboodian 1969). The third method, quantile regression, has been used to fit distributions to data before, but has seldom been applied Gaussian mixtures. We introduce a straightforward method of implementing quantile regression to fit a Gaussian mixture to data. We also compare the methods and describe the advantages and disadvantages of each method.

There are also ways to optimize a multivariate Gaussian mixture, which is necessary for the Black-Litterman model. In this paper, we only detail the method of Maximum Likelihood estimation as it is usually sufficient for the purposes.

We then introduce the Mean Variance Optimization Model and the Black-Litterman Model. Finally, we present a revised Black-Litterman model which incorporates Gaussian mixtures.

2 Gaussian Mixtures

When a data set cannot be accurately modeled with a single normal distribution, it is useful to use a Gaussian Mixture instead. Theoretically, Gaussian mixtures can be used to model any continuous distribution to within a non-zero error bound so long as we allow for an arbitrarily large number of components in the Gaussian mixture.

The density function of a Gaussian mixture is as follows:

$$f_{GM}(x) = \sum_{i=1}^{m} \lambda_i \cdot f_i(x), \qquad f_i \sim \mathcal{N}(\mu_i, \sigma_i), \qquad \sum_{i=1}^{n} \lambda_i = 1 \tag{1}$$

Where λ_i is the probability of a data point being picked from normal distribution i.

The multivariate case is similar:

$$f_{GM}(x_1, x_2, ..., x_k) = \sum_{i=1}^{m} \lambda_i \cdot f_i(x_1, x_2, ..., x_k), \qquad f_i \sim \mathcal{N}_k(\mu_i, \Sigma_i) \tag{2}$$

where $u_i = (u_{i1}, u_{i2}, ..., u_{in})^T$

2.1 Optimizing a Univariate Gaussian Mixture

Several methods exist for fitting a univariate Gaussian mixture to data. However, it is not always clear which method ought to be used. Since optimality can be measured in several different ways, one method might seem preferable over another in one regard, yet also seem lacking in another regard. The primary difference between the existing methods of fitting a Gaussian mixture to data is that they each define optimality differently. That being said, each method has its own unique benefits and drawbacks, and so none should be ruled out entirely.

2.1.1 Method of Moments

The Method of Moments (Pearson 1894) is a straightforward yet somewhat tedious way to find a two-component Gaussian mixture which matches the first five

moments of a data set.

The formula for the k_{th} moment of a distribution is

$$M_k = E(X^m) \tag{3}$$

And given a particular data set, we can calculate the sample moments as

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{4}$$

A two-component univariate Gaussian mixture has the parameters μ_1, μ_2, σ_1, σ_2, λ_1, and λ_2, but since λ_2 can be rewritten as $1 - \lambda_1$, the mixture be described by only five parameters. Formulas for the first five moments a Gaussian mixture in terms of it's components can be obtained rather easily. These moments can then be equated with the moments of the data to produce a system of five equations. Naturally, the system can be solved (through a great deal of algebraic manipulation) to obtain the values of the five parameters.

The following algorithm was originally presented in Cohen, 1967: [3].

Consider the following nonic equation:

$$\sum_{i=0}^{9} a_i z^i = 0 \tag{5}$$

where

$a_9 = 24$ $a_7 = 84k_4$ $a_5 = 90k_4^2 + 72m_3k_5$ $a_3 = 288m_3^4 - 108m_3k_4k_5 + 27k_4^3$ $a_1 = -96m_3^4k_4$
$a_8 = 0$ $a_6 = 36m_3^2$ $a_4 = 444m_3^2k_4 - 18k_5^2$ $a_2 = -(63m_3^2k_4^2 + 72m_3^3k_5)$ $a_0 = -24m_3^6$

and,

$$k_4 = m_4 - 3m_2^2, \quad k_5 = m_5 - 10m_2m_3$$

Let z_0 be a suitable negative real root of equation 5. Define r, d_1, d_2 as follows:

$$r = \frac{-8m_3z_0^3 + 3k_5z_0^2 + 6m_3k_4z_0 + 2m_3^3}{z_0(2z_0^3 3k_4z_0 + 4m_3^2)} \quad d_1 = \frac{r - sqrt(r^2 - 4z_0)}{2} \quad d_2 = \frac{r + sqrt(r^2 - 4z_0)}{2}$$

Finally, we have explicit equations for our parameters:

$$\hat{p} = \frac{d_2}{d_2 - d_1} \tag{6}$$

$$\hat{\mu}_1 = d_1 + \bar{x} \tag{7}$$

$$\hat{\mu}_2 = d_x + \bar{x} \tag{8}$$

$$\hat{\sigma}_1^2 = \frac{d_1(2r - m_3)}{3z_0} + m_2 - d_1^2 \tag{9}$$

$$\hat{\sigma}_2^2 = \frac{d_2(2r - m_3)}{3z_0} + m_2 - d_2^2 \tag{10}$$

2.1.2 Maximum Likelihood

Perhaps the most popular method for fitting a Gaussian mixture to data is the maximum likelihood method. This involves a likelihood function, which is a probability function which takes in the parameters of the data.
This function is defined as:

$$L(\Theta|X) = P(X|\Theta) = \prod_{i=1}^{N} p(x_i|\Theta) \tag{11}$$

Where X represents the data, and Θ represents the set of parameters. The goal is to maximize L given the data. Typically $log(L(\Theta|X))$ is maximized instead because it is easier, yet still provides the same result. [12].

For a Gaussian mixture there is not a direct way to maximize $L(\Theta|X)$, so the Expectation Maximization (EM) algorithm is often used to approximate the solution. Here we will briefly discuss the univariate case, and we will discuss the multivariate case in the next section. The univariate case is described by Jim Wang, 2000 [16], and previously by Behboodian (1969) [2].

The general idea behind the EM algorithm is to assume the data, X, is incomplete, and to generate more data from the distribution $f(x)$. The idea follows such that one uses estimated parameters as true parameters, and then they use estimated values as observed values. The estimated parameters are used to guess missing data,

and the estimated values are used to guess the missing parameters. This process repeats until convergence.

In the case of mixed distributions, the simplest way to think of this "missing data" is to say the values tell us which Gaussian corresponds to which data item.

Assumption: Given M normal distributions, $y_i \in 1, ..., K$ and y_i=k if the i^{th} data point was generated by the k^{th} Gaussian.

Your data, $x_1, ..., x_s$ is treated as a random sample from 1.

Define:

$$\hat{w}_{ij}^{m} = \frac{p_j(x_i; \mu_j^{(m)}, \sigma_j^{(m)})}{m \sum_{l=1}^{k} p_l(x_i; \mu_l^{(m)}, \sigma_l^{(m)})} \tag{12}$$

where i is the number of data points ranging from $1, ..., s$ and j is the number of Gaussians ranging from $1, ..., k$.

$\hat{w}_{ij}$ is a weighted probability of each data point for each normal distribution. One can intuitively see that the higher $\hat{w}_{ij}$ is, the more likely that the data point belongs to the j^{th} normal distribution.

This ultimately fills the "missing data" piece of the EM algorithm. The other half of the EM algorithm and results of Behoodian(1969) are:

$$a_j^{(m+1)} = \sum_{i=1}^{s} \hat{p}_j^{(m)} \hat{w}_{ij}^{(m)} \tag{13}$$

$$\hat{\mu}_j^{(m+1)} = \sum_{i=1}^{s} \hat{w}_i j^{(m)} x_i \tag{14}$$

$$\hat{\sigma}_j^{2(m+1)} = \sum_{i=1}^{s} \hat{w}_{ij}^{(m)} x_i^2 - (\hat{(\mu)}_j^{(m)})^2 \tag{15}$$

where j=1,...,k and m is the number of the iteration. The process is straight forward from here. Take the initial guesses $p_j^{(0)}$, $\mu_j^{(0)}$, $\sigma_j^{(0)}$ and compute $\hat{w}_{ij}^{(0)}$. Then plug $\hat{w}_{ij}^{(0)}$ in to find your parameters, and repeat. This iteration process will eventually converge to a local maximum.

For derivations of these formulas and more information on the Maximum Likelihood method see Behboodian [2], Blimes [12], and Wang [16].

2.1.3 Quantile Regression

Quantile regression is another method of fitting a univariate Gaussian mixture to data. The idea is to minimize the sum of the squared distances between the mixed-Gaussian quantiles and the data quantiles. Essentially, it is a non-linear constrained least squares optimization of the quantiles of the data.

Though it may seem intuitive to use optimization to solve for each the quantiles of a Gaussian mixture, this is computationally costly. Rather, it is more effective to calculate the quantiles directly. Since an explicit quantile function for Gaussian mixtures does not exist, we use reverse cubic spline interpolation to create an approximate quantile function for a Gaussian mixture. Using a cubic spline to calculate the quantiles is generally one to two orders of magnitude faster than using optimization to solve the CDF for each quantile.

Definition a *cubic spline curve* is a piecewise cubic curve with a continuous second derivative.

One would use a cubic spline if they had points and wanted to create a line connecting them. However in our case we want to find the points (quantiles) being connected in the CDF, which is a reverse cubic spline process. We used the built-in MATLAB function spline (interchanging the roles of x and y) to accomplish this, however a breakdown of reverse cubic spline interpolation can be found in [5].

The first step of the reverse cubic spline interpolation entails calculating the value of the CDF at a linearly spaced vector of x values. Once the function values at these points are determined, a cubic spline can be created with cumulative probability as the independent variable and the x value as the dependent variable. This cubic spline will approximate the quantile function for the Gaussian mixture.

The accuracy of the cubic spline is highly dependent on the initial choice for the range of the x values and the number of total points the CDF is calculated at. To set the bounds of the data points, it is generally a good idea to take the smallest

mean, subtract some multiple of its associated standard deviation and set that as a lower bound. Likewise, to create an upper bound, one should take the largest mean and add some multiple of its associated standard deviation.

In the tables below, we display the sum of the absolute value of the residuals between the actual quantile and the quantiles calculated by a spline created from 1000, 320 ($\approx 10^{5/2}$), 100, and 32 ($\approx 10^{3/2}$) points. We also show how this value changes as the range of x values changes. The first table shows these values when calculating the standard 99 quantiles at .01, .02, ..., .99. The second table shows these values for the calculation of 999 quantiles at .001, .002, ..., .999.

Table 1: 100 Quantiles

	3σ	4σ	5σ
1000 pts	8.1375e-09	2.1283e-08	4.3482e-08
320 pts	8.7968e-07	1.4274e-06	3.4156e-06
100 pts	6.7359e-05	2.2255e-04	4.1812e-04
32 pts	2.3759e-02	6.0522e-02	3.9589e-01

Table 2: 1000 Quantiles

	3σ	4σ	5σ
1000 pts	1.3555e-01	2.2240e-07	5.0257e-07
320 pts	1.5330e-01	2.2235e-05	4.6372e-05
100 pts	2.1917e-01	2.2785e-03	4.6902e-03
32 pts	6.7066e-01	6.1045e-01	4.2743e+00

Clearly, if any of the quantiles are close to 0 or 1, the interval should be larger, yet using a larger range while keeping the number of points the same will at some point produce less accurate estimations since the points will become more spread out and produce a less accurate fit. We find that using 100 points and three standard deviations for the bounds will be sufficient in most cases, but if any of the desired quantiles are close to 0 or 1, slightly larger bounds may produce better predictions. As we can see in table 4, using more than 4 standard deviations on the bounds will rarely prove to be beneficial even when calculating quantiles close to 0 or 1. In general, using more points will increase the accuracy.

Adding additional points increases run time as well though. Below is a table

Table 3: Run Time

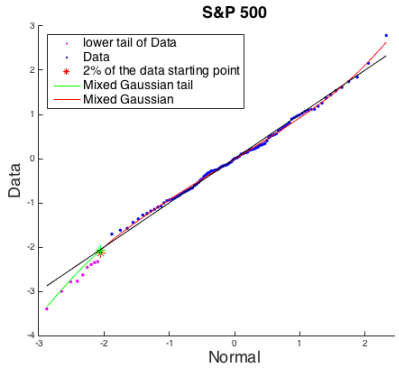
Number of points	1000	320	100	32
Run time	0.228118s	0.060493s	0.011330s	0.004360s

with the run times in MATLAB for the calculation of the standard 100 quantiles with 1000, 320, 100, and 32 points used to create the cubic spline. The mount of run time calculating additional quantiles adds is negligible since creating the cubic spline requires comparatively much more computation than calculating the value of the spline at different quantiles does.

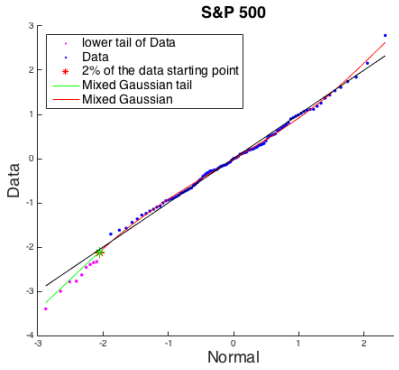
By doing the least squares optimization, run time becomes a more important issue since finding a local minimum may require calculating the quantiles of a Gaussian mixture hundreds of times and doing a global optimization will require even more iterations than before.

2.2 Method Comparison

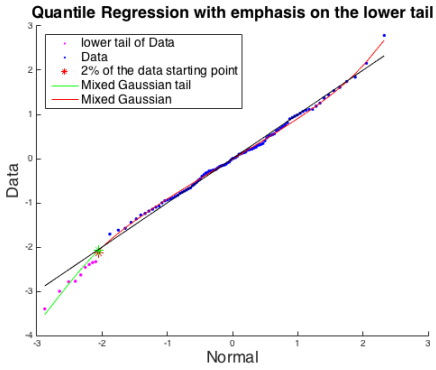
Each method has benefits and drawbacks, so it is wrong to say that any particular method should be used in all situations. Maximum Likelihood is often the preferred method for fitting Gaussian mixtures to data because it usually produces a reasonably good fit in a short amount of time. One drawback of MLE is that it will only converge to a local maximum. This can be partially remediated by repeating the EM algorithm for multiple starting points, but there is no guarantee of finding the global maximum. Similarly, quantile regression will only converge to a local minimum, but it has the added benefit that the quantiles can be chosen to fit the data as desired. The Method of Moments is preferable in that it will fit the first moments exactly and thus eliminates the problem of local minimums.We can see from all three graphs below that each method fits the data fairly well, so the decision of which method to use will depend on the user's preference.



Maximum Likelihood



Method of Moments



Quantile Regression

2.3 Multivariate Mixtures

Multivariate Gaussian mixtures are useful in the case where one is dealing with multiple variables which codepend on each other in some way. When the higher comoments of a distibution do not coincide with the higher comoments of a multivariate normal distribution, a multivariate Gaussian mixture can be used instead. As a reminder, the formula for a multivariate Gaussian mixture is as follows:

$$f_{GM}(x_1, x_2, ..., x_k) = \sum_{i=1}^{m} \lambda_i \cdot f_i(x_1, x_2, ..., x_k), \qquad f_i \sim \mathcal{N}_k(\mu_i, \Sigma_i) \tag{16}$$

where $u_i = (u_{i1}, u_{i2}, ..., u_{in})^T$ and λ_i is the probability of a data point being picked from the i^{th} multivariate normal distribution.

The equations for the overall mean vector and covariance matrix of a multivariate Gaussian mixture are as follows.

$$\mu = \sum_{i=1}^{m} \lambda_i \mu_i \tag{17}$$

$$\Sigma = \Sigma_{i=1}^{m} \lambda_i \Sigma_i + \Sigma_{i=1}^{m} \lambda_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{18}$$

The intuition behind 17 ought to be clear, but we present a proof of 18 in the

appendix. The latter equation proves to be quite significant in the derivation of the revised Black-Litterman model.

2.3.1 Multivariate Maximum Likelihood Method

A popular method used to optimize the parameters for a multivariate Gaussian mixture is the multivariate Maximum Likelihood method, specifically the EM Algorithm.

To refresh the general idea displayed in the univariate section: when using the EM Algorithm for mixed densities, we consider our data set, X, incomplete and assume unobserved data, Y, exists such that the values y_i indicate which density corresponds to which data item [12].

In this case, the densities are different multivariate normal distributions.

The general mixed-density maximum likelihood function is computed using Bayes rule (Bilmes 1998) [12]:

$$Q(\Theta, \Theta^{(p-1)}) = \sum_{l=1}^{m}\sum_{i=1}^{n} \log(a_l) p(l|x_i, \Theta^{(p-1)}) + \sum_{l=1}^{m}\sum_{l=1}^{n} \log(p_l(x_i|\theta_l)) p(l|x_i, \Theta^{(p-1)}) \tag{19}$$

Where m is the number of probability density functions, m is the number of data points, and a_l is the probability that a given data point will come from the l^{th} density function. Note that $\sum_{l=1} a_i = 1$. $\Theta = [a_i, ...a_m, \theta_1, ...\theta_m]$ is the set of parameters, where $\theta = (\mu_x, \Sigma_x)$.

Next the maximum likelihood function is calculated by plugging a multivariate Gaussian mixture pdf into an equation similar to 19, where μ_x is a vector of means ($\mu_x = (\mu_{x1}, ..., \mu_{xk})$) and Σ_x is a $k \,\mathrm{x}\, k$ covariance matrix.

$$Q(\Theta, \Theta^{p-1}) = \sum_{l=1}^{m}\sum_{i=1}^{n} r_{il} log(a_l) - \frac{1}{2}\sum_{l=1}^{m}\sum_{i=1}^{n} r_{il} log(2\Pi|(\Sigma_{x|l}^{-1}|) - \frac{1}{2}\sum_{l=1}^{m}\sum_{i=1}^{n} r_{il}(x-\mu_{x|l})\Sigma_{x|l}^{-1}(x-\mu_{x|l})^T \tag{20}$$

where r_{il} is $p(y_i = l|x_i, \Theta)$ which is computed using Bayes rule:

$$p(y_i = l|x_i, \Theta) = \frac{p(y_i = l)p(x_i|y_i = l, \Theta_l)}{p(x_i|\Theta)}$$

$$r_{il} = \frac{a_l p_l(x_i|\Theta_l)}{\sum_{l'=1}^{L} a_{l'} p_{l'}(x_i|\Theta_l)} \tag{21}$$

The subscript $x|l$ indicates the parameter is an estimate of X conditioned on the Gaussian mixture component l.

By taking the partial derivatives of $Q(\Theta, \Theta^{p-1})$ in terms of the parameters, one obtains the following equations:

$$\hat{\mu}_{x|l} = \frac{\sum_{i=1}^{N} r_{il} x_i}{\sum_{i=1}^{N} r_{il}}$$

$$\hat{\Sigma}_{x|l} = \frac{\sum_{i=1}^{N} r_{il}(x_i - \mu_{x|l})(x_i - \mu_{x|l})^T}{\sum_{i=1}^{N} r_{il}}$$

One begins with initial guesses for the parameter Θ to find r_{il}, which is then used to calculate $\hat{\mu}_{x|l}$ which is then used to find $\hat{\Sigma}_{x|l}$. $\hat{\mu}_{x|l}$ and $\hat{\Sigma}_{x|l}$ are then used to find a new r_{il}. This iterative process is then repeated until the stopping criteria are met.

3 Incorporating Gaussian Mixtures into the Black-Litterman Model

3.1 The Models

In 1952, Markowitz introduced the Mean-Variance Optimization Model (MVOM) as an approach for optimizing portfolio weights. This model takes into account the expected excess returns (μ) and risks (Σ) of the assets in order to do one of the following:

1. Maximize the returns on a portfolio, given some level of risk
2. Minimize the risk on a portfolio, given some value of returns

In 1992, Black and Litterman developed a portfolio allocation model that allows an investor to incorporate views about the market and obtain a modified expected return vector and covariance matrix. This model assumes two distinct sources of information: the market equilibrium and the investor views. Using the Bayesian Framework, the BLM combines these two sources of information to define a new expected return vector $\bar{\mu}$ and covariance matrix $\bar{\Sigma}$ which can then be used with the MVOM to find the optimal portfolio weights.

The Black-Litterman model assumes asset returns are normally distributed, but it is well known that asset returns often do not follow a normal distribution. In actuality, the tail behavior of the distributions tends to be quite different from a normal distribution, with the returns often having heavier tails and skewness which a normal distribution cannot account for. Consequently, we assume the returns follow a Gaussian mixture distribution. Furthermore, we use the method proposed by Martellini and Ziemann (2007) to calculate the expected return vector (Π) based on a utility function of four moments rather than the two-moment utility function used in the original Black-Litterman model.

3.2 Higher Moment Utility Functions

As with the original Black-Litterman approach, we do not wish to use the historical returns to predict future returns. However, unlike Black and Litterman, we will use a higher moment version of the CAPM. The MVOM assumes a utility function of 2 moments:

$$U(w) = w^T\mu - \delta w^T\Sigma w \tag{22}$$

Where: μ is a column vector of the expected asset returns
Σ is the covariance matrix of the returns
w is a column vector of the portfolio weights

Martellini and Ziemann instead assume a utility function of 4 moments:

$$U(w) = w^T\mu - \delta_1 w^T\Sigma w + \delta_2 w^T S w \otimes w - \delta_3 w' K w \otimes w \otimes w \tag{23}$$

Where $\otimes$ is the Kronecker product, μ, Σ, and w are defined as before, and S and K represent the coskewness and cokurtosis matrices of the returns. Martellini and Ziemann also propose a method for determining the coefficients δ_1, δ_2, and δ_3. Since the regular Black-Litterman model cannot account for the higher order comoments of the data, such a utility function cannot be used. Our model is useful in that it can accounts for the higher comoments of the data and so a higher moment utility function such as 23 can be used.

3.3 Revised Black-Litterman Model

In the original Black-Litterman model, the pdf of the equilibrium was given by

$$f(x_1, ..., x_n) = \frac{1}{(2\pi)^k|\Sigma|}\exp(-\frac{1}{2}(x-\Pi)^T\Sigma^{-1}(x-\Pi)) \tag{24}$$

and the pdf of the mean was given by

$$m(x_1, ..., x_n) = \frac{1}{(2\pi)^k|\Sigma|}\exp(-\frac{1}{2\tau}(\Pi - E[r])^T\Sigma^{-1}(\Pi - E[r])) \tag{25}$$

Using a Gaussian mixture, the pdf will now be given by

$$f_{MG}(x_1, ..., x_n) = \sum_{i=1}^{m} \frac{\lambda_i}{(2\pi)^k|\Sigma_i|}\exp(-\frac{1}{2}(x-\Pi_i)^T\Sigma_i^{-1}(x-\Pi_i)) \tag{26}$$

with the pdf of the mean given by

$$m_{MG}(x_1, ..., x_n) = \sum_{i=1}^{m} \frac{\lambda_i}{(2\pi)^k|\Sigma_i|}\exp(-\frac{1}{2\tau}(\Pi_i - E[r])^T\Sigma_i^{-1}(\Pi_i - E[r])) \tag{27}$$

The covariance matrices can be found using a multivariate Gaussian mixture fitting method. The equilibrium mean vectors $\Pi_1, ..., \Pi_m$ must be obtained separately though, and so we present an intuitive and straightforward way of doing this.

When fitting the distribution based on the historical returns, the mean vectors for each component $(\mu_1, ..., \mu_m)$ were determined by optimizing the Gaussian mixture and the overall first moment of the distribution, μ, was given by equation 17. We now desire the overall first moment to be Π. Equation 18 indicates another factor we ought to take into account. Altering the distances between the means of the individual components will have an effect on the overall covariance of the distribution. We present the following method to accommodate for this:

We define

$$\Pi_i = \Pi + (\mu_i - \mu) \tag{28}$$

Theorem 3.1. *Let g be a multivariate Gaussian mixture with mean vectors $\mu_1, ..., \mu_m$ and covariance matrices $\Sigma_1, ..., \Sigma_m$. If the means are redefined as in equation 28, the first moment of the Gaussian mixture will equal Π, yet the distribution will have the same overall shape as g.*

Proof. First we will show that the overall first moment is Π:

$$\begin{aligned}
E_i[E[\Pi_i|i]] &= \sum_{i=1}^{m} \lambda_i \Pi_i \\
&= \sum_{i=1}^{m} \lambda_i (\Pi + (\mu_i - \mu)) \\
&= \sum_{i=1}^{m} \lambda_i \Pi + \sum_{i=1}^{m} \lambda_i \mu_i - \sum_{i=1}^{m} \lambda_i \mu \\
&= \Pi \sum_{i=1}^{m} \lambda_i + \mu - \mu \sum_{i=1}^{m} \lambda_i \\
&= \Pi + \mu - \mu \\
&= \Pi
\end{aligned}$$

Also, we can see that the distance between means still remains the same:

$$\Pi_j - \Pi_k = \Pi + \mu_j - \mu - (\Pi + \mu_k - \mu) = \mu_j - \mu_k$$

Since the distances between the means are preserved and the covariance matrices remain the same, the shape of the Gaussian mixture will remain the same.

□

Using Gaussian mixtures in the Black-Litterman model also necessitates a change in the view expectation vector. The view matrix, P, can remain the same and can be used for each component of the Gaussian mixture, since the same assets are involved in the view regardless of which mean vector we are considering. However, the view expectation vector, Q, cannot remain the same. If we were to use the same Q for each component, this would have an adverse effect. To illustrate this, we present the following simple example for a single stock:

	μ	σ	λ
$\mathcal{N}_1$	1	1	.5
$\mathcal{N}_2$	2	1	.5

If we were to introduce a view that the return on the stock will be two percent and then implement the Bayesian framework on each component, μ_1 would not change at all since it already has a value of 2, but μ_2 would change drastically since it is relatively far from 2. It seems more intuitive for the components to be impacted in a simlar way (though not necessarily identically). In order to shift the components more similarly, we create an expectation vector Q_i for each component of the Gaussian mixture in such a way that the overall expectation of the view will still be Q.https://www.sharelatex.com/project/55ad18aa9896abcb1f2705c5

$$Q_i = Q + P(\mu_i - \mu) \tag{29}$$

Theorem 3.2. *If Q and $\mu_1, ..., \mu_n$ are vectors of length n and we define $Q_1, ...Q_n$ as in equation 29, then $E_i[E[Q_i|i]] = Q$*

Proof.

$$\begin{aligned} E_i[E[Q_i|i]] &= \sum_{i=1}^{m} \lambda_i Q_i \\ &= \sum_{i=1}^{m} [\lambda_i Q + \lambda_i P(\mu_i - \mu)] \\ &= Q \sum_{i=1}^{m} \lambda_i + P \sum_{i=1}^{m} \lambda_i \mu_i - P\mu \sum_{i=1}^{m} \lambda_i \\ &= Q + P\mu - P\mu \\ &= Q \end{aligned}$$

For the covariance matrices of the views, There are multiple ways of defining each Ω_i. If a desired covariance matrix is specified, one can see from equation 18 that if we define Ω_i by

$$\Omega_i = \Omega - (Q_i - Q)(Q_i - Q)^T \tag{30}$$

the overall variance will be Ω. However, we propose a modification of the method defined by He and Litterman (1999) to define each Ω_i. Since we already have each covariance matrix Σ_i, we can define each Ω_i as a diagonal matrix where the diagonal elements are the vector $P\Sigma_i P^T \tau$.

$$\Omega_i = \begin{bmatrix} p_1 \Sigma_i p_1^T \cdot \tau & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_n \Sigma_i p_n^T \cdot \tau \end{bmatrix} \tag{31}$$

4 Conclusion

The Primary Benefit of this revised Black-Litterman model is that it allows for the preservation of the data whereas the original Black-Litterman only preserves the second comoment of the data. The primary usefulness of this is that it allows for more accurate data simulation and also

4.1 Further Research

References

[1] Black, Fischer and Litterman, Robert (1992) "Global Portfolio Optimization.", *Financial Analysts Journal*, 48(5):28–43.

[2] Behboodian,J.,1970, "On a mixture of normal distributions" *Biometrika*, 57,215-217.

[3] Cohen, A.C., 1967, "Estimation in mixtures of two normal distributions," *Technometrika*, 9,15-28.

[4] Bertsimas, Dimitris, Vishal Gupta, Ioannis Ch Paschalidis. 2012. Inverse optimization: a new perspective on the black-litterman model. *Operations research* **60**(6) 1389–1403.

[5] Boor, C. (2001). A Practical Guide to Splines: Applied Mathematical Sciences (Rev. ed.). New York: Springer.

[6] A. P. Dempster, N. M. Laird and D. B. Rubin. Maximum Likelihood from Imcomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society.* Series B (Methodological) Vol. 39, No. 1 (1977), pp. 1-38

[7] Harvey, C. R., LiechtyJ, C., LiechtyM, W. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10, 469–485b

[8] He and Litterman, Robert (1999), "The Intuition Behind Black-Litterman Model Portfolios", 1999. Goldman Sachs Asset Management Working paper.

[9] Idzorek, Thomas (2005), "A Step-By-Step guide to the Black-Litterman Model, Incorporating User- Specified Confidence Levels", Working paper.

[10] Markowitz, H. (1952), PORTFOLIO SELECTION. *The Journal of Finance*, 7: 77–91

[11] The EM Algorithm. Statistics 580. University of California, Los Angeles Lecture Notes

[12] J. Bilmes. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden markov models. Technical Report ICSI-TR-97-02, University of Berkeley, 1997 .

[13] T. Hastie, R. Tibshirani, and J. H. Friedman. The Elements of Statistical Learning. Springer, 2001

[14] G. J. McLachlan and K. E. Basford. Mixture Models: Inference and Applications to Cluster- ing. Marcel Dekker, Inc., New York, Basel, 1988

[15] G. J. McLachlan and D. Peel. Finite Mixture Models. Wiley Series in Probability and Statistics. Wiley, New York, 2000

[16] J. Wang, Modeling and generating daily changes in market variables using a multivariate mixture of normal distributions, Working paper. Department of Math and Computer Science, Valdosta State University, January 2000

5 Appendix

Proof.

$$
\begin{aligned}
Cov[X] &= E[(X - E(X))(X - E(X))^T] \\
&= E[(X - E(X))(X^T - E(X)^T)] \\
&= E[XX^T] - E[XE(X)^T] + E[E(X)E(X)^T] + E[E(X)E(X)^T] \\
&= E[XX^T] - E[X]E[X]^T - E[X]E[X]^T + E[X]E[X]^T \\
&= E[XX^T] - E[X]E[X]^T \\
&= E_i[E[XX^T \mid i]] - E_i[E[X \mid i]]E_i[E[X \mid i]]^T \quad \text{(Law of Total Expectation)} \\
&= E_i[Cov(X) \mid i] + E_i[E[X \mid i]E[X \mid i]^T] - E_i[E[X \mid i]]E_i[E[X \mid i]]^T \\
&= E_i[Cov(X) \mid i] + Cov_i(E[X \mid i]) \\
&= \Sigma_{i=1}^m \lambda_i \Sigma_i + \Sigma_{i=1}^m \lambda_i (\mu_i - \mu)(\mu_i - \mu)^T
\end{aligned}
$$

□

Proof.

$$
\begin{aligned}
&\lambda_1(\mu_1-\mu)(\mu_1-\mu)^T+\lambda_2(\mu_2-\mu)(\mu_2-\mu)^T\\
=&\lambda_1(\mu_1-\lambda_1\mu_1-\lambda_2\mu_2)(\mu_1-\lambda_1\mu_1-\lambda_2\mu_2)^T+\lambda_2(\mu_2-\lambda_1\mu_1-\lambda_2\mu_2)(\mu_2-\lambda_1\mu_1-\lambda_2\mu_2)^T\\
=&\lambda_1((1-\lambda_1)\mu_1-\lambda_2\mu_2)((1-\lambda_1)\mu_1-\lambda_2\mu_2)^T+\lambda_2((1-\lambda_2)\mu_2-\lambda_1\mu_1)((1-\lambda_2)\mu_2-\lambda_1\mu_1)^T\\
=&\lambda_1(\lambda_2\mu_1-\lambda_2\mu_2)(\lambda_2\mu_1-\lambda_2\mu_2)^T+\lambda_2(\lambda_1\mu_2-\lambda_1\mu_1)(\lambda_1\mu_2-\lambda_1\mu_1)^T\\
=&\lambda_1\lambda_2^2(\mu_1-\mu_2)(\mu_1-\mu_2)^T+\lambda_2\lambda_1^2(\mu_2-\mu_1)(\mu_2-\mu_1)^T\\
=&\lambda_1\lambda_2^2(\mu_1-\mu_2)(\mu_1-\mu_2)^T+\lambda_2\lambda_1^2(\mu_2-\mu_1)(\mu_2-\mu_1)^T\\
=&(\lambda_1\lambda_2^2+\lambda_2\lambda_1^2)(\mu_1-\mu_2)(\mu_1-\mu_2)^T\\
=&(\lambda_2+\lambda_1)(\lambda_1\lambda_2)(\mu_1-\mu_2)(\mu_1-\mu_2)^T\\
=&\lambda_1\lambda_2(\mu_1-\mu_2)(\mu_1-\mu_2)^T
\end{aligned}
$$

□

□