A probabilistic fuzzy approach to modeling stock price behavior

Dennis Bron

February, 2006

Master thesis Informatics & Economics
School of Economics, Erasmus University Rotterdam
Supervisors: Dr. ir. Jan van den Berg, Ludo Waltman MSc.

Abstract

In this thesis, a probabilistic fuzzy model that produces a conditional probability for a stock return given a lagged return, is fitted on historical stock returns. The model consists of three rules: one for "low", one for "medium", and one for "high" lagged returns. For each fuzzy class of lagged returns, a conditional probability distribution is approximated using a mixture model. The three rules are interpolated using a fuzzy reasoning scheme, similar to that of Takagi-Sugeno fuzzy inference systems. Using the resulting model, potential stock price paths can be simulated. From these price paths we can deduce an option price that is valid for risk-neutral investors that are not able to benefit from arbitrage opportunities.

Acknowledgements
I would like to express my gratitude to my supervisors for this thesis. Without Jan van den Berg and Ludo Waltman, writing it would never have been such a joyful and also time-efficient experience. Furthermore, I would like to thank Eelco Koppert for providing me with the data, Alexander Sprengers and Pepijn Füss for the lively and useful discussions, and my roommates Henno Talsma, Friso Ypma for providing the necessary distractions from time to time. Last but not least, I would like to thank my parents for sponsoring my study: a great part of my life that hereby comes to an end.

Index

1. Introduction

1.1 General background

In this thesis, I will investigate the impact of the conclusions of a paper by Jan van den Berg, Uzay Kaymak and Willem-Max van den Bergh, all professors at the Erasmus University in Rotterdam. In "Financial markets analysis using a probabilistic fuzzy modeling approach", they find that short-term stock returns are clearly not independent, but show conditional probabilities for reversal that are significantly higher than the unconditional probabilities for the same returns. To get to these conclusions, they use a probabilistic fuzzy approach, in which two types of uncertainty are modeled, namely fuzzy (linguistic) and probabilistic uncertainty.

They have built a probabilistic fuzzy system (PFS) that classifies returns as "very low", "low", "medium", "high" or "very high". Given the fuzzy classification of the last return, the model produces the conditional probabilities for the following return.

The key advantage of a PFS is that its rules can be interpreted, while interpretability is clearly not one of the strengths of the most widely used methods for option valuation. All the additions to the Black-Scholes model that are made over the past decades to correct for volatility smiles and conditional heteroskedasticity have made an already reasonably complex model into one that can only be understood by real specialists.

Another advantage of using a probabilistic fuzzy method is the model-free approach. Little assumptions about stock-price behavior have to be made.

1.2 Research goal

The key question to be answered by this research can be stated as follows:

Does a probabilistic fuzzy approach based on intraday data produce output probability density functions that are significantly different for different lagged returns?

1.3 Methodology

In this thesis, a probabilistic fuzzy system with mixture models describing the output conditional probability density functions is estimated from historical stock returns. The input space (the lagged return) is partitioned by a certain number of predetermined triangular membership functions. The parameters of the mixture models describing the pdf for the following return are optimized using the maximum likelihood method. The model could be viewed as being semi-parametric: part of the model structure is predetermined; part of it is induced from the data. This approach is aimed at achieving the best of both worlds: avoid assuming a function that is incapable of forming a good representation of the underlying process, as can happen when using parametric models, and at the same time limit the amount of parameters that can grow very quickly when using non-parametric models (Bishop (1995)).

To test whether this model is capable of capturing the patterns that I expect to be present in the stock return data, I will first fit the model on an artificial dataset that contains a high conditional probability for 'reversal'.

After that, I will fit the model to a time series of one-minute returns and later on five-minute returns. From the resulting pdf's, conclusions can be drawn about the amount of 'reversal' and conditional variance in the data.

Using these approximated conditional probability distributions; it is possible to value short-lived options by simulating possible price paths and computing the expected value of the option at maturity.

1.4 Relevance

Stock option pricing is a very relevant research topic, given the enormous volumes of options that are traded around the world every day, the amount of money involved, and the known anomalies within existing models.

Furthermore, the analysis of intraday data is still in its infancy. Lack of computational power and storage capacity until recently seriously limited its use. That day-trading and intraday stock returns are very relevant are possibly best illustrated by the fact that many traders in the world's stock markets are not even allowed to hold overnight positions for their own account.

1.5 Unique aspects

The following aspects of the research conducted in this thesis are, as far as I know, unique:

- Practical application of probabilistic fuzzy systems with output probability density functions modeled by mixture models existing of multiple Gaussians;
- Monte Carlo simulation using semi-parametrically estimated conditional probability distributions estimated by probabilistic fuzzy systems with output pdf's modeled by mixture models.

1.6 Structure of the thesis

In the next chapter, a further introduction into the topic of option pricing and modeling stock price behavior is given. First the basics of options and option markets are explained, followed by two of the most popular models for option valuation. The chapter is concluded by stating my assumptions about the stock pricing process, the impact this has on the expected results of my experiments, and how I use these to compute option prices.

In chapter three, the probabilistic fuzzy approach is further explained, followed by the method I use to optimize the parameters.

The implementation of the model is illustrated in chapter four. I will explain which part of the model structure is preset, which parameters are chosen for the optimization, and how the data is prepared.

In chapter five I will show the results of the various experiments and in chapter six I will elaborate further on the impact of these findings.

2. Stock option valuation

2.1 Option basics

An option contract gives the right to buy or sell an underlying asset for a given price at some point in the future. Options that give the right to buy are called call options; options that give the right to sell are called put options. The predetermined price at which the transaction is executed is called the strike price. European options can only be executed at the expiration date, whereas American style options can be executed by the contract holder at any time on or before the expiration date.

The value of an option contract at the time of the execution (for European style options always the expiration date) is determined by the difference between the price of the underlying asset and the strike price of the option contract. Before expiration, it is harder to determine the value of the contract, because it depends on the expected price of the asset at expiration and potentially even a risk premium.

In this thesis, only European call options are considered. The price of put options can be deduced from call option prices using the put-call parity (see Appendix).

Looking at American style options using probabilistic fuzzy models would be very interesting, but stretches beyond the scope of this thesis.

2.2 Option markets

The introduction of stock options on an organized exchange dates back to 1973. Since then the volumes in option markets have grown dramatically. Options are traded on many exchanges in the world and underlying assets range from stocks and stock indices to foreign exchange rates, futures, and even stock price volatility. This thesis is concentrated only on the valuation of stocks and – very similar – stock indices. These kinds of options are also the most common and most widely traded.

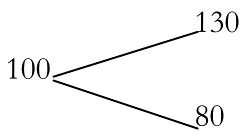
2.3 Option valuation models

In general, models for option valuation take the following six factors on which option prices depend into account:

1. The current stock price S_t;
2. The strike price X;
3. The time to expiration T;
4. The volatility of the stock price during the option's life σ_T;
5. The risk free interest rate during the option's life r;
6. The dividends expected during the option's life d.

Binomial models

In 1979 Cox, Ross, and Rubinstein introduced binomial trees as a simple analytical method for option valuation. Consider a call option on a stock currently worth 100, with the chance to be worth either 80 or 130 at expiration:



The exercise price of the option contract is 100, so the pay-off of the option is either 30 or 0. Now, consider a portfolio consisting of a short position in the option, i.e. you have 'written' or sold the option, and a long position in part of the share, Δ. The portfolio at expiration is worth either Δ130 – 30 or Δ80 + 0. It is possible to choose Δ so that the portfolio is riskless, or the pay-off is equal in both cases.

$$\Delta 130 - 30 = \Delta 80 + 0 \Rightarrow \Delta 50 = 30 \Rightarrow \Delta = 0.6$$

At expiration of the option, the portfolio is worth $0.6 \cdot 80 - 0 = 48$, or equivalently $0.6 \cdot 130 - 30 = 48$. As the current value of Δ stocks (60) is known, the current value of the option can be computed if no-arbitrage is assumed:

$48e^{-rT} - c = 60 \Rightarrow c = 60 - 48e^{-rT}$.

In general terms the binomial model can be summarized as follows:

Consider a stock whose initial price is S_0 and an option initially worth f. Suppose that the option's life is T and that during this time the stock price can either move up to S_0u, or down to S_0d (where $u > 1$ and $d < 1$). The option's payoff when the stock prices moves up is f_u, and the option's payoff when the stock price moves down is f_d.

Again we have a portfolio consisting of a long position in Δ shares and a short position in one option. We calculate Δ to make the portfolio riskless.

$$S_0u\Delta - f_u = S_0d\Delta - f_d \Rightarrow \Delta = \frac{f_u - f_d}{S_0u - S_0d}$$

Discounted at the risk free interest rate, the present value is

$\left(S_0u\Delta - f_u\right)e^{-rT}$ (or equivalently $\left(S_0d\Delta - f_d\right)e^{-rT}$).

The cost of setting up the portfolio is $S_0\Delta - f$ and must be equal to the expected payoff discounted at the risk free rate. Substituting for the equation for Δ, the price of a call option is:

$$f = e^{-rT}\left[p\, f_u + (1-p)\, f_d\right]$$

Where

$$p = \frac{e^{rT} - d}{u - d}$$

This option pricing formula above does not take the stock's expected return into account. It does not matter if the chance of the stock price moving up is 90% or

50%. The reason is that we are not valuing the option in absolute terms, but in terms of the current price of the underlying asset. The probabilities of the stock price moving up and down are already incorporated in the current stock price.

Of course in the real world, stock prices do not move either up or down to known values. A commonly used way to use binomial models however is to build trees that exist of different layers of up and down movements with the standard deviation of the stock price returns for that interval length. This creates a large number of potential price paths that has the same standard deviation as the stock returns. According to the central limit theorem (Appendix), option prices from a binomial model are equivalent to Black-Scholes option prices if the number of layers in the tree is sufficiently large.

The Black-Scholes Model

In the early 1970s Fischer Black, and Myron Scholes developed a framework for option valuation. Robert Merton offered the mathematical proof some years laters. The resulting Black-Scholes model proved to be a revolution is option valuation and earned them a Nobel price. In spite of its obvious shortcomings, the Black-Scholes model is still widely used today.

To price derivatives that rely on the value of an underlying asset, it is necessary to describe the process that the price of the underlying asset will follow in the future. The value depends heavily on the nature of the stochastic process followed by the asset price. An example of such a process is a geometric Brownian motion, on which the famous Black-Scholes option-pricing model is based.

Brownian motion

In the Black-Scholes world, the stock prices are generated by a generalized Wiener process called a geometric Brownian motion (Hull (2003)) that describes the price change dS in terms of a constant drift μ_S of the stock, the standard deviation σ_S of the stock, a period of time dt, and a stochastic term γ, which is a drawing from a standard normal distribution.

$$dS = \mu_S dt + \sigma_S \gamma \sqrt{dt}$$

This kind of process is often referred as a 'random walk' because it has no structure or statistical properties to give traders the opportunities of having an expected return other than μ_S, or the risk free rate plus a risk premium that depends on σ_S. This property of the Brownian motion is consistent with the Efficient Market Hypothesis (EMH), see Hull (2003).

Assumptions

The following assumptions are made when deriving the Black-Scholes-Merton differential equation:

1. the stock price follows a geometric Brownian motion with μ and σ constant;
2. short selling of the asset is allowed;
3. there are no transaction costs or taxes;
4. the asset is perfectly divisible;
5. there are no dividends during the life of the derivative;
6. there are no riskless arbitrage opportunities;
7. asset trading is continuous;
8. the risk-free rate of interest, r_f, is constant and the same for all maturities.

Assumption six means that every riskless asset should earn the same return, namely the risk-free interest rate r_f. Assumption five can be relaxed.

Using these assumptions it is now possible to derive the Black-Scholes model (see Hull (2003)).

$$c = S_0 N(d_1) - X\, e^{-r(T-t)} N(d_2) \ ,$$

Where c is the price of a call option, S_0 is the initial stock price, X the strike price of the option, $N(x)$ the cumulative standard normal density function and where

$$d_1 = \frac{\ln\left(S_0 / X\right) + \left(r + \sigma_T^2 / 2\right)\left(T - t\right)}{\sigma_T \sqrt{\left(T - t\right)}}$$

$$d_2 = \frac{\ln\left(S_0 / X\right) + \left(r - \sigma_T^2 / 2\right)\left(T - t\right)}{\sigma_T \sqrt{\left(T - t\right)}} = d_1 - \sigma_T \sqrt{\left(T - t\right)}$$

2.4 The stock pricing process

In 1986, French and Roll have published an article that was aimed at explaining the role of publicly available information – news, in the stock pricing process. They acknowledge three different sources of volatility in stock prices:

1. Public information;
2. Private information;
3. Trading noise or mispricing.

Their conclusion is that, although there is a significant amount of trading noise present in stock returns, the majority of the short term volatility is caused by public information becoming available.

They come to this conclusion by looking at hourly returns and the difference in volatility between trading hours and non-trading hours.

Using realized volatilities for volatility estimation when pricing options has gained popularity over the past decade. Realized volatilities are calculated using intraday returns and should according to widely accepted financial theory deliver an error-free estimation of the interday volatility (Andersen and Bollerslev (1998)). When ticker level returns (typically one-minute) are used however, daily volatilities are structurally overestimated. Corsi et al. (2001) show that one-minute returns are about twice as volatile as daily returns and can be scaled to produce unbiased volatility estimations. They fail to propose an explanation for the observed phenomenon.

It seems likely that the additional volatility in short term returns is caused by trading noise.

If $R_0, R_1, \cdots, R_T$ are independent random drawings from an identical distribution, then

$$\mathrm{Var}\left(\sum_{t=0}^{T} R_t\right) = \sum_{t=0}^{T} \mathrm{Var}\left(R_t\right).$$

In Corsi et al. (2001) however,

$$\mathrm{Var}\left(\sum_{t=0}^{T} R_t\right) \ll \sum_{t=0}^{T} \mathrm{Var}\left(R_t\right).$$

Therefore $R_0, R_1, \cdots, R_T$ cannot be independent drawings from an identical distribution as often assumed in financial econometrics.

In addition, one would expect that pricing errors will be corrected over time when the market's consensus expectations arise from trading. That is, until newly available information alters the market's expectations. This would imply that short-term stock returns are not independent.

If we consider a 'real' stock price and an observed one, blurred by the influence of trading noise, one would expect that for instance a very low short-term return is very likely to be followed by a reversal, as a very low return is likely to be composed of a negative 'real' return and a negative pricing error. As the pricing error will be corrected over time, the probability for an, at least partly, reversal is larger than the unconditional probabilities for the same behavior.

This hypothesis also offers a new way of looking at the phenomenon 'overreaction and reversal' often studied in behavioral finance (Fama (1998)). Given an extremely negative or positive return, one would expect either a very large price change caused by newly available information, a great mistake in the pricing process, or a combination of both. This leads to a conditional probability distribution of the following return that will differ from the unconditional returns that are typical for a Markov process. Please note that I am talking about conditional probability: I do not state that traders overreact to news in general, but I do state that extreme returns might coincide with overreaction rather than underreaction.

Another reason for assuming a pricing process with conditional probabilities for lagged returns that differ from the unconditional probabilities lies in the modeling of conditional volatilities. Incorporating conditional volatilities is an adjustment that is often made to the Black-Scholes-Merton-world by practitioners because of the obvious amount of heteroskedasticity present in real life stock returns. It is not hard to identify periods of significantly low and high volatility in most financial time series.
Conditional pdf's offer a way to deal with conditional volatilities, although the precondition for the conditional volatility in this thesis (the lagged return) is different from that in well-known models like GARCH (Bollerslev (1986)), where the future volatility depends on the historic volatility, which is never directly observed.

The third reason for using a probabilistic fuzzy approach to modeling stock price behavior is that the structure of the model is very straightforward and easy to understand. Many of the option pricing models used in practice have grown to very complex versions of the Black-Scholes model, incorporating conditional heteroskedasticity (Bollerslev (1986)), stock-price jumps, volatility jumps, and corrections for volatility smiles (Derman and Kani (1994). The rules of the general form

$$\text{If } R_{t-1} \text{ is } A_j \text{ then } p(R_t) = \varphi(R_t, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma})$$

Where A_j is a fuzzy class, $p(R_t)$ the pdf of R_t, the return a time t, and $\varphi(R_t, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ a mixture model depending on parameters vectors $\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}$, should be interpretable for anyone that has, at least a moderate, understanding of the Black-Scholes model.

In the following chapter I will explain how I will estimate the conditional probability distributions from historical stock price data. Next, I will use the Monte Carlo method to numerically estimate an expected payoff from the found conditional pdf's.

2.5 Assuming an unobserved consensus price and pricing errors

In the efficient market hypothesis (Hull (2003)), stock-prices are equal to the market's consensus estimate of the sum of the future cash flows discounted at the risk-free rate. This consensus estimate is a result of many traders that are trying to maximize their profits.

The actual consensus stock price cannot be observed until all market participants have declared their orders, consisting of a number of stock and a limit price at which they are willing to sell or buy. Some traders however, will wait for the orders of others before making their own decisions. To create a more efficient market, the market maker who is in charge of the trade in a given asset will publish bid and ask prices. These are the lowest price at which he is willing to sell for his own account and the highest price at which he is willing to buy. Of course these prices are based on the market maker's estimate of the consensus expectations. If there is no trade, he will have a hard time estimating and the spread between bid and ask prices will be large. This creates a type of volatility in stock returns that is sometimes called the bid-ask-bounce. Usually the bid-ask bounce is only taken into consideration when dealing with data on a transaction level, which is even more precise than one-minute stock returns.

If we look at stock returns, we can assume the stock returns to be composed of changes in consensus estimates and changes in the pricing error.

$$R_{obs,t} = \frac{dS_{obs,t}}{S_{obs,t-1}} = \frac{dS_{cons,t} + d\varepsilon_t}{S_{obs,t-1}} = \frac{dS_{cons,t} + d\varepsilon_t}{S_{cons,t-1} + \varepsilon_{t-1}}$$

Where $R_{obs,t}$ is the observed stock return at time t, $S_{cons,t}$ the market's consensus stock price, $S_{obs,t}$ the observed stock price, ε_t the pricing error, and $dX_t = X_t - X_{t-1}$.

We would expect the changes in the market's consensus expectations $dS_{cons,t}$ to be irregular shocks, caused by newly available public or private information. The pricing error is likely to be high after a shock, especially when caused by private information, and will tend to zero over time as transactions occur. Pricing error can also be caused by errors in the trading process itself. On Friday December 9, 2005 for instance, one of Japan's major brokers Mizuho Securities Co. sold 610,000 shares at 1 yen instead of 1 share at 610,000 yen. This simple typing error cost the firm at least 27 billion yen (225 million dollar) and had great influence on the stock price for a short period of time.

Assuming and unobserved consensus price and pricing errors that tend to be corrected over time leads to the following hypotheses about the conditional distribution of returns for "low", "medium" and "high" lagged returns:

1. For "low" and "high" lagged returns, the mean of the conditional probability distribution will shift slightly in the opposite direction ('reversal') when looking at the right time-frame;
2. The conditional distribution of returns, given a "low" or "high" lagged return will show a larger volatility than the one given "medium" lagged returns.

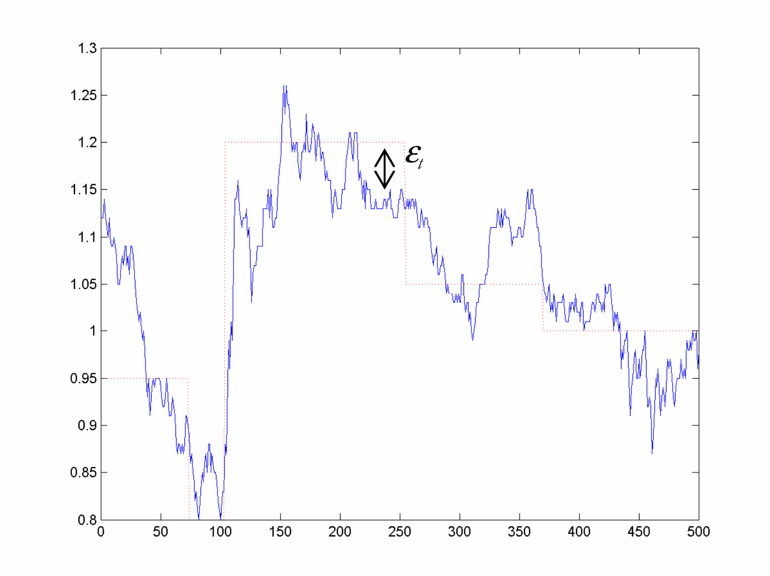


Figure 2.1 - An observed stock price (solid line) and imaginary consensus estimates (dotted line)

2.6 The option price

In Duffie (1996), it is proven that if it is assumed that there are no-arbitrage opportunities, each asset value should equal its expected value discounted at the risk-free interest rate. This argument is called risk-neutral valuation, because it cancels out any risk premiums. In markets that are not completely efficient however, there could be arbitrage opportunities.

Assuming pricing errors that tend to be corrected over time does not rule out the presence of arbitrage opportunities. These opportunities will be very small however, as taking advantage of them will quickly restore the equilibrium prices.

In our experiments, we will determine the option's expected payoff. This expected payoff, discounted at the risk-free rate gives the option price for a risk-neutral investor that is not able to benefit from arbitrage opportunities:

$$c_0 = e^{-rT} E\left(c_T\right) = e^{-rT} E\left(\min\left(0, S_T - X\right)\right).$$

Here c_0 is the option price at time 0, c_T is the option price at time T, S_T is the stock price at time T, and X is the strike price of the option.

3. A probabilistic fuzzy approach

3.1 Probabilistic fuzzy systems

Probabilistic fuzzy systems were introduced in 2002 by Jan van den Berg and Uzay Kaymak. These models combine fuzzy inference systems with probabilistic mathematics and are therefore capable of dealing with two types of uncertainty:

1. The fuzzy nature of human language (input, reasoning);
2. Probabilistic uncertainty.

In their paper "Financial markets analysis by using a probabilistic fuzzy modelling approach", they show that PFS are very well suited for capturing skewed conditional probability distributions in financial time series.

In "On probabilistic connections of fuzzy systems" Van den Berg and Kaymak (2004) introduce a model that combines fuzzy uncertainty with probabilistic uncertainty. Probabilistic fuzzy systems are for instance able to answer a simple question like: "What is the change that a random Dutch woman is tall?". To answer this question, two questions have to be answered and combined:

1. To what degree is a woman of a given length called "tall"? (fuzzy uncertainty)
2. What is the probability distribution of the variable length over the population of Dutch women? (probabilistic uncertainty)

To answer the question, a mapping is made from an input space (a woman's length and the associated membership of the linguistic concept "tall") to an output space (the probability distribution of Dutch women being tall) using a number of rules and a certain reasoning scheme to make interpolations between these rules.

3.2 Fuzzy partitioning and normalized memberships

A fuzzy partitioning is a set of *j* membership functions that span the entire input or output space of a fuzzy inference system, where for each value of *x* a the sum of all membership value is 1.

$$\sum_{j} \mu_{A_j}(x) = 1, \forall x \in X$$

A fuzzy partitioning is a useful and intuitive way of ensuring that all the probabilities in a probabilistic fuzzy system sum up to 1. If probabilities are assigned to all classes A_j, these probabilities can simply be multiplied with the membership values. However, this can also be realized by using normalized membership values.

$$\bar{\mu}_{A_j}(\mathbf{x}) = \frac{\mu_{A_j}(\mathbf{x})}{\sum_{j'=1}^{a} \mu_{A_{j'}}(\mathbf{x})}$$

3.3 Conditional probabilities

Let (x_1, y_1), …, (x_n, y_n) denote a random sample of size n. Using this sample, an estimate of $\Pr(C|A)$, the conditional probability of event C given event A is provided by the following statistical formula

$$\hat{p}(C \mid A) = \frac{\sum_{i=1}^{n} \chi_A(x_i)\chi_C(y_i)}{\sum_{i=1}^{n} \chi_A(x_i)},$$

Where the characteristic functions χ_A and χ_C are given by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

And

$$\chi_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{otherwise} \end{cases}$$

Now suppose that A and C are fuzzy events instead of ordinary crisp events. This means that A and C are defined by the membership functions μ_A and μ_C instead of the characteristic functions χ_A and χ_C.

$$\hat{p}(C \mid A) = \frac{\sum_{i=1}^{n} \mu_A(x_i)\mu_C(y_i)}{\sum_{i=1}^{n} \mu_A(x_i)} .$$

This formula is based on Zadeh's definition of the probability of a a fuzzy event (Zadeh (1968)) and can be used for estimating the probability parameters in a probabilistic fuzzy system.

3.4 Fuzzy histograms

The probability distribution over the output space of probabilistic fuzzy systems can be approximated using the fuzzy histogram approach as described in Van den Bergh, Kaymak, and Van den Berg (2002).

Traditional histograms have been used for a long time as a simple method to approximate and visualize probability density functions. Observations are divided into bins of a certain length. These bins are plotted against the number of observations per bin to give an impression of the probability distribution.

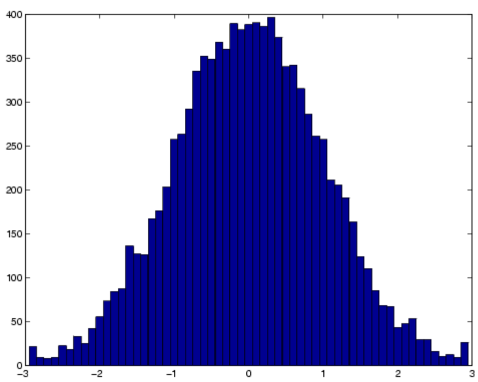


Figure 3.1 - Example of a crisp histogram

Obviously the rigidity of this division into bins is a serious drawback of regular histograms when trying to approximate a probability distribution for simulations. All approximated probability density functions are shaped like a staircase, which is quite an unusual shape in real life. Ludo Waltman (2005) proves that a fuzzy histogram, in which an observation belongs to a certain bin to a certain degree, always gives the better approximation of the underlying probability distribution at the very least if the number of observations goes to infinity and the size of the bins to infinitely small.

The probability density function *f(y)* can be approximated using a fuzzy histogram consisting of fuzzy columns

$$\hat{f}_j(y) = \frac{\Pr(C_j)\mu_j(y)}{\int_{-\infty}^{\infty}\mu_j(y)\,dy}$$

Where $\int_{-\infty}^{\infty}\mu_j(y)\,dy$ is a scaling factor representing the fuzzified size of C_j. Summing all the colums $\hat{f}_j(y)$ will approximate the complete probability density function $\hat{f}(y)$.

$$\hat{f}(y) = \sum_j f_j(y)$$

3.5 A simple regression using a probabilistic fuzzy system

The concepts in paragraphs 3.2, 3.3 and 3.4 have led to probabilistic fuzzy systems as proposed in Kaymak, Van den Bergh, Van den Berg (2003). In chapter 6 of Waltman (2005) however, it is shown that such a Probabilistic Mamdani Fuzzy System, performs very badly when trying to estimate a simple linear function with a normally distributed error term of the form:

$$f(x) = ax + b + N(\mu, \sigma)$$

Ludo Waltman's explanation for the poor performance of PFS on such a simple problem is the method of interpolation between two pdf's in the output space. The interpolation is done by averaging the membership values of the different membership functions in the output space weighted by their respective firing rate of the associated rule. This results in an estimation of the pdf that is severely biased towards dilution.

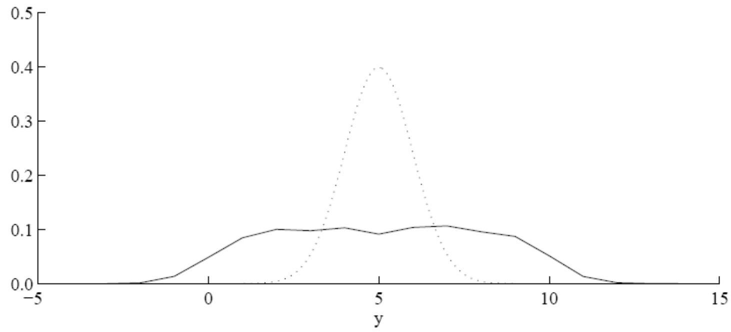


Figure 3.2 - A diluted estimation (solid line) of the underlying pdf (dotted), from Waltman (2005)

A possible solution is to consider a function $\phi(y;\alpha_{j1},...,\alpha_{jm})$, where α_{jm} denotes parameter *m* for rule *j*, that describes the probability density function, resulting in a fuzzy system with rules of the following general form:

If $\mathbf{x}$ is A_j then $p(y)=\phi(y;\alpha_{j1},...,\alpha_{jm})$

Where $\int \phi(y;\alpha_{j1},...,\alpha_{jm})dy=1$ and $\phi(y;\alpha_{j1},...,\alpha_{jm})\geq 0, \forall y$ and $\phi(y;\alpha_{j1},...,\alpha_{jm})$ is the same function for every rule. Now we can interpolate the parameters to produce an estimate of the pdf.

$$\alpha_m(x)=\sum_j \overline{\mu}_{A_j}(x)\alpha_{jm}$$

Ludo Waltman shows that if one chooses the normal distribution for $\phi(y;\alpha_{j1},...,\alpha_{jm})$ and an input space partitioned by triangular membership functions, one can get, not very surprisingly, an excellent estimation of a linear regression with normally distributed error terms. He proposes to use mixture models for more complex applications.

3.6 Optimizing the parameters using the maximum likelihood criterion

To optimize the parameters of the mixture models describing the conditional pdf's for each of the rules in our model, the maximum likelihood criterion is used. The basic principle of maximizing the likelihood criterion is choosing the model parameters such that it maximizes the chance of having the observed sample data for the entire space of possible models:

$$m^* = \arg\max_{m \in M} \Pr\left(D \mid m\right)$$

Where *m** is the optimal model given the data, *M* is the model space, and *D* is the dataset available for training.

The likelihood of a dataset consisting of data pairs (*x*, *y*) is given by:

$$L\left(\mathbf{x}, \mathbf{y}\right) = \prod_i \Pr\left(y_i \mid x_i\right)$$

As the number of instances grows, the likelihood of the dataset will approach zero. This inevitably leads to computational problems. To solve these problems, the loglikelihood method is often used.

Maximizing the natural logarithm of the likelihood is equivalent to maximizing the likelihood, as the natural logarithm is a continuously increasing function. Taking the logarithm of a product of probabilities is equivalent to summing the logarithms of these probabilities, and this will be much easier to implement.

$$\log L\left(\mathbf{x}, \mathbf{y}\right) = \ln \prod_i \Pr\left(y_i \mid x_i\right) = \sum_i \ln \Pr\left(y_i \mid x_i\right)$$

3.7 Mixture models

A mixture model is a weighted linear combination of a number of Gaussian pdf's.

$$y = \sum_i w_i \cdot \frac{1}{\sigma_i \sqrt{2\Pi}} \cdot e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}}$$

The parameters of the mixture models to be optimized are $\mathbf{w}, \boldsymbol{\sigma}, \boldsymbol{\mu}$: the vectors of weights, standard deviation and means of the Gaussians. This means that the interpolation of the parameters is done as follows:

$$w_m(x) = \sum_j \overline{\mu}_{A_j}(x) w_{jm}, \ \mu_m(x) = \sum_j \overline{\mu}_{A_j}(x) \mu_{jm} \text{ and } \sigma_m(x) = \sum_j \overline{\mu}_{A_j}(x) \sigma_{jm}.$$

When using mixture models as the output for PFS, we have to make sure that the right parameters of one mixture model are interpolated with the right parameter of the other. If the maximum likelihood method is used to optimize the parameters for all rules at once however, we elegantly avoid this problem as the interpolation itself is part of the optimization.

3.8 Using an evolution strategy to maximize the likelihood criterion

To optimize the parameters of the fuzzy system using the maximum likelihood criterion, I use an Evolution Strategy. Evolution Strategies are based on the biological concept of evolution. A population of a certain size is simulated during a certain number of generations. Each generation, parents with a high 'fitness' create offspring that is subject to mutations. If the offspring is generated by combining parameters from multiple parents, we call this 'cross-over'. Mutation is vital for an Evolution Strategy to prevent the algorithm from getting trapped in local optima.

To decrease the tendency to get trapped in a local minimum or maximum further, we use tournament selection to produce the parents for each new generation. By randomly selecting pairs from the existing population and using only the winner of each pair in terms of fitness, every individual (except the very worst performing one) has a chance to be selected, but the selection is biased towards the better performing individuals.

In this thesis, the fitness of each individual is measured by its likelihood.

One of the problems when using evolutionary methods for fitting models is choosing the appropriate standard deviation for mutation. Mutation works as follows:

$$x = x + \mathrm{N}(0, \sigma)$$

with N(μ, σ) a drawing from the standard normal distribution with mean μ and standard deviation σ. When σ is too small, it takes too long before the optimum is reached, when σ is too large the mutations can consistently be too large to find the optimum. Schwefel (1974) has solved this problem by introducing self-adaptation, where σ itself is also subject to mutation. In this way, the σ will gradually decrease as the population reaches the optimum.

3.9 Monte Carlo simulation

The Monte Carlo method was invented in 1946 by Stanislaw Ulam, a Polish born mathematician who worked in the United States, while trying to compute the probability of winning a game of solitaire. It is a very intuitive and widely used method. Its major drawback is its computational expensiveness. It is easy to code Monte Carlo simulations that run for days, even on modern computer systems. Coming up with an analytical alternative however, can take even longer and, in many cases, is not even possible.

Monte Carlo simulation, as it is understood today, involves using statistical sampling to approximate solutions to a quantitative problem. Looking back at the Brownian motion in the Black-Scholes framework the Monte Carlo method is easily explained.

$$dS = \mu_S dt + \sigma_S \gamma \sqrt{dt}$$

It is possible to generate the random variable γ using a pseudo-random number generator and compute *dS* for every step in the time-path, leading to a certain stock price at expiration of the option. If you repeat this experiment many times, a good estimation of the probability distribution of the option's payoff can be made.

To reduce the number of computational operations needed with the Monte Carlo method, a number of techniques have been developed. In summary, they all involve searching the output space in a more or less structured way, reducing the estimation error when approximation the pdf. These methods are called Quasi Monte Carlo (QMC).

In principle one can view the multi-layered binomial trees mentioned before as a very simplistic way of Quasi Monte Carlo simulation.

In Lemieux and l'Ecuyer (2001) a clear overview of the use of QMC methods in finance is presented. With Monte Carlo simulation, we try to estimate a probability density function by choosing a point set from the function's domain. These points are independent and usually generated by a pseudo random number generator.
The idea of QMC is to use a more regularly distributed point set that is typically deterministic in stead of stochastic. This will effectively reduce the standard error of the estimation of the underlying probability distribution.

3.10 Generating pseudorandom number from the estimated distributions

In order to create pseudorandom number that are distributed according to our estimation of the conditional probability distribution, we need to convert pseudorandom number that are uniformly distributed from 0 to 1. Uniformly distributed pseudorandom numbers are available in each programming language and on every computer. In order to do this, we need to compute the cumulative probability distribution of our estimated pdf.

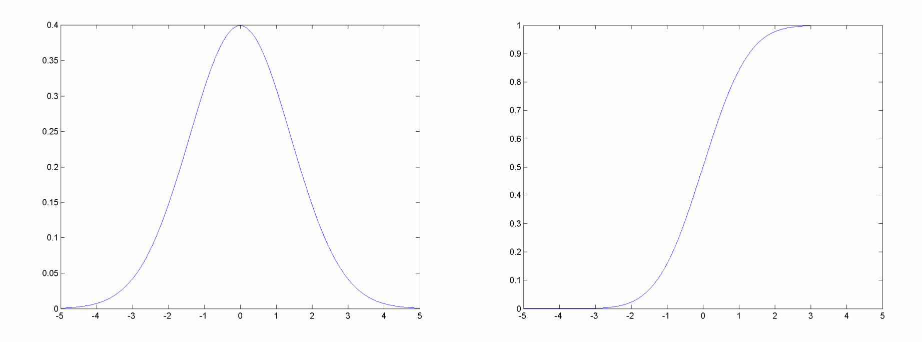


Figure 3.3 - A standard normal pdf (left) and the associated cumulative probability distribution (right)

The cumulative distribution associated with a pdf is the sum of all probabilities lower than the number under consideration, or in mathematical terms:

$$\mathrm{cdf}\left(x\right) = \int_{-\infty}^{x} \mathrm{pdf}\,(x)dx$$

As the cumulative distribution by definition ranges from 0 to 1, we have a useful mapping from the uniform distributed numbers available in the programming language to the empirical pdf. If $r_{uniform}$ is a pseudo-random variable, uniformly distributed between 0 and 1, then a random variable distributed according to the estimated pdf is computed by:

$$r = \mathrm{cdf}^{\mathrm{inv}}\left(r_{uniform}\right).$$

To compute the inverse cumulative distribution, we need to numerically approximate the integral of the pdf up to the point where it equals $r_{uniform}$ and then take that point as our new random variable. In pseudo-code:

$$r_{uniform} \in \mathrm{Uniform}\left[0,1\right]; y = -y_{limit}; p_{cum} = 0;$$
$$\text{while } p_{cum} < r_{uniform} \text{ and } y < y_{limit}$$
$$\{ \quad p_{cum} = p_{cum} + dy \cdot p\left(y \mid x\right);$$
$$y = y + dy; \}$$
$$r = y - dy;$$

4. Experimental setup

4.1 Choosing the model structure

Before the parameters of the model can be optimized, a number of choices regarding the structure of the model have to be made. For this reason, this approach can be viewed as being 'semi-parametric' (Bishop (1995)). These choices could also be made using an evolutionary algorithm making the model entirely 'non-parametric', but this would lead to even longer computation times and almost certainly to overfitting of the data. Because the maximum likelihood is always realized by a probability distribution showing a peak on every data point and zero probability elsewhere, a certain generalization has to be made. In this case, I choose to limit the amount of Gaussians the mixture model exists of and the minimum standard deviation of each Gaussian component to prevent peaks on incidental clusters of data points.

The maximum likelihood always has the tendency to overfit the data. Consider e.g. a mixture model existing of two Gaussians and normally distributed data. The maximum likelihood will always be for a normal distribution with one small peak on the two data points that happen to be closest together.

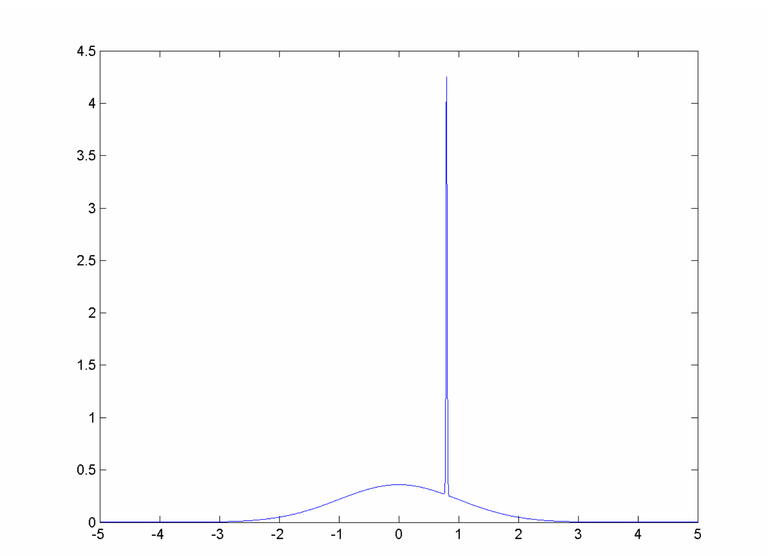


Figure 4.1 - A mixture model with 2 Gaussians on normally distributed data, showing one peak on an incidental data cluster

Besides choosing the number of parameters for the mixture models describing the output space, I will have to choose the number of rules considered and the width of the fuzzy concepts spanning the input space.

I choose to use only three rules: one for "low" lagged returns, one for "medium" and one for high". Between the three resulting probability distributions a linear interpolation of their parameters is made because I divide the input space with triangular membership functions. The left and right membership functions have their peak at R_{limit}, this point is chosen so that almost every data point falls within the range of the membership function under consideration. In some cases, it might be that for one or two outliers the distance beyond R_{limit} is ignored. This leads to a better approximation of the conditional pdf's than including all data points because stretching the input space to contain all outliers would mean that we have less data available for the rules associated with "low" and "high" returns.

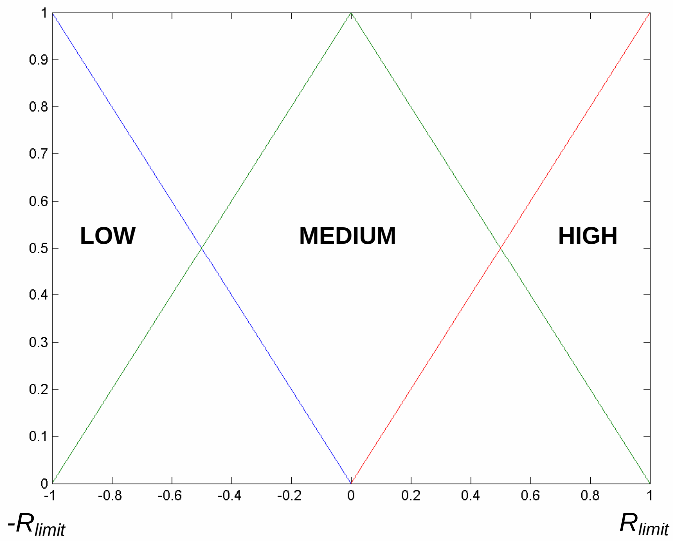


Figure 4.2 - Partitioning of the input space

4.2 Optimizing the parameters

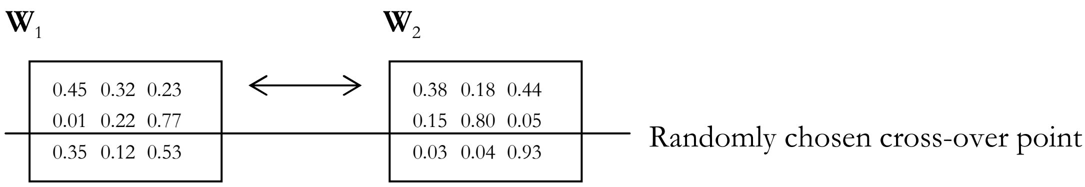
All mixture models are described by three vectors: **w**, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. All individuals in the evolution strategy are composed of a set of rules; each associated with a mixture model, and can be described by parameter matrices **W**, **M**, **S**.
In order to optimize the parameters, the parameter matrices first have to be initialized for all individuals of the first generation.

Initialization
For the matrix **W**, all values are drawn from a uniform distribution between 0 and 1 and then divided by the sum of all values in their respective row. This is to ensure that the weights for each mixture model sum to 1.

For **M**, all values are drawn from a normal distribution with mean 0 and standard deviation $R_{limit}/10$.

For **S**, all values are initialized at 1 because I use normalized data and I quickly found out that this could help the algorithm in quickly finding the optimum and decrease the tendency to get stuck in local optima.

Cross-over
The parents are selected using tournament selection as described in paragraph 3.10. Cross-over is implemented as a single point cross-over, where the two parents interchange a number of parameter vectors, up to a random row in the parameter matrices. For each parameter matrix a new row is chosen.
Cross-over probability is set at 0.7.



Mutation
Mutation is implemented including self-adaptation, which means that each individual has its own associated standard deviation σ (not to be confused with

the vector of standard deviations of the Gaussian components), determining the width of the probability distribution from which the mutations are drawn. The σ 's are mutated using $\sigma_\sigma = \sigma/5$, however each σ must remain larger than or equal to 0.01.

The mutation probability is set at 0.5.

If a parameter matrix is mutated, random drawings from the normal distribution with mean 0 and standard deviation σ are added. To ensure that the weights for each mixture model still sum to 1, they are again divided by the sum of their respective row. The weights have to remain at least 0.001.

The standard deviations of the Gaussian components of the mixture models have to remain at least 0.1 to prevent overfitting.

Elitism
To prevent the maximum likelihood of the population from downfalls over time, the best 5 individuals of the previous generation are added to the next generation to replace random individuals at the end of each generation.

Number of generations and population size
The algorithm will run for a predefined number of generations an population size. Increasing both will lead to higher likelihoods until the global optimum is found, but both at a heavy computational prize.

4.3 Preparing the data

I will use one-minute stock quotes from the Microsoft stock (see Appendix), downloaded using HQuotes (www.hquotes.com).

Date, time	Open	High	Low	Close	Volume
22-4-2005 14:10:00	25.12	25.13	25.11	25.13	94489
22-4-2005 14:11:00	25.12	25.14	25.12	25.14	62077
…	…	…	…	…	…

For every minute there is an opening, highest, lowest and closing quote. I take the average of open and close to approximate the average quote. Using these average quotes, I compute one-minute returns. I choose not to look at the highest and lowest quotes because these could have been for very small transactions and therefore have less meaning than the opening and closing quotes.

Because the first return, recorded at the opening of the stock market at 9:30 AM is not a one-minute return but an overnight return, I choose to delete these from my data to estimate the model on one-minute returns only. Because I cannot use the first return as lagged returns too, the first return I can actually use is from 9:32 AM.

Next I divide all the returns by their standard deviation because the model is set up to work best for standard deviations around 1. One-minute return standard deviations are of course far from 1, usually less than 0.1%. I do not subtract the mean because 0 seems to be a very common value for one-minute returns, even though there are no returns without associated trade volume present in the data. I would like this to be shown explicitly in the estimated probability distributions as 0, and not as minus the mean divided by the standard deviation.

Logreturns

In financial econometrics, usually logreturns are used when examing financial time series. There are three reasons for this. The first reason is that if one assumes that the stock price follows a geometric Brownian motion,

$$dS = \mu S dt + \sigma S dz$$

Then

$$d \ln S = \left(\mu - \sigma^2/2\right) dt + \sigma dz \, .$$

This means that

$$\ln S_T - \ln S_0 = \ln \frac{S_T}{S_0} \sim N\left[\left(\mu - \sigma^2/2\right)T, \sigma\sqrt{T}\right],$$

Where S_T is the stock price at a future time T, S_0 is the stock price at time 0, and $N(m, s)$ denotes a normal distribution with mean m and standard deviation s.

This means that in a geometric Brownian motion, the logreturns $\ln \frac{S_T}{S_0}$ are normally distributed.

The second reason is that consecutive logreturns can simply be summed to produce the total return. In this way it is much easier to remain precise when implementing a simulation that consists of a series of return, then when the total return is computed by multiplying a large number of returns.

The third reason is that if we draw logreturns from a given probability distribution, it is impossible to accidentally produce negative stock prices. This is of course crucial in simulation as a negative stock price is impossible in reality.

I do not make the assumption that stock returns are lognormally distributed, but I do use logreturns to prevent negative stock prices and to be able to sum consecutive logreturns.

4.4 Simulating stock price paths

In paragraph 3.10, I have explained how we can draw pseudorandom variables that are distributed according to the distributions given by the probabilistic fuzzy model. To get from these returns to an estimate of the expected value of the option, we need to simulate many potential price paths. For each resulting stock price, we subtract the strike price of the option and compute the resulting option value. From these potential option values, we can compute the expected value, equalling the option price for a risk-neutral investor.

$$c_0 = e^{-rT} E\left(c_T\right) = e^{-rT} E\left(\min\left(0, S_T - X\right)\right)$$

Where c_0 is the current option price, c_T is the option value at expiration, S_T is the stock price at expiration and X the strike price of the option.

The expected value is better estimated with more potential price paths. The best way to implement the simulation is by first calculating a lookup table with the next return for all potential lagged returns and all possible values of a random number uniformly distributed from 0 to 1, with a limited level of precision. Such a lookup table would enable quick simulation of potential price paths. Calculating it however, is very computationally expensive. For e.g. 1,000 possible lagged returns and 1,000 possible random numbers, 1,000,000 returns have to be computed. To compute each of these numbers, first the integral of a pdf, resulting from interpolating multiple mixture models, has to be numerically approximated. This approximation can take up to a second. Calculating a lookup table with one million values would take about a week on a high-end pc.

For the purpose of this thesis, I will simply simulate a large number of potential price paths. I will use the model I estimated for five-minute returns and 100 consecutive returns, resulting in 500 minute price paths. The distribution of the resulting 500 minute returns can be compared with the theoretical Black-Scholes distribution and with the distribution of 500 minute returns in the actual data.

5. Results

5.1 A simple experiment

To test whether the model explained in the previous chapter works in practice, we will conduct a simple experiment. I have created a dataset that incorporates a very clear pattern of 'reversal': the mean of the conditional probability distribution of R_t shifts R_{t-1} to the opposite direction.

$$R_t = \mathrm{N}(0,1) - R_{t-1}$$

Where N(0, 1) is random drawing from the standard normal distribution.

The dataset contained 1001 data pairs where R_{t-1} ranged from -1 to 1 with step size 0.002. The input space was partitioned using overlapping triangular membership functions. Triangular membership functions are used because I want to make a linear interpolation between the output pdf's of the different rules. Only a limited number of membership functions and associated rules need to be considered as the range of the input variable is known (from -1 to 1).

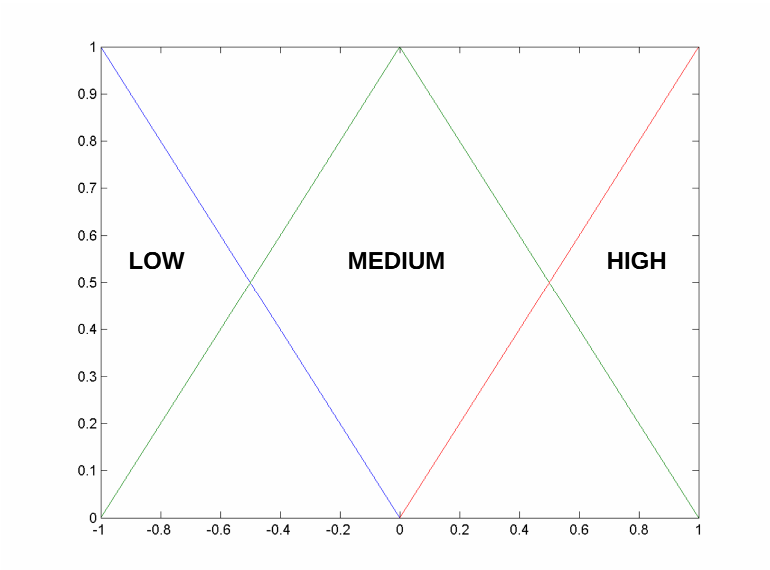


Figure 5.1 - The partitioning of the input space using 3 triangular membership functions

These three membership functions each have one associated rule of the form

If R_{t-1} is A_j then $p(R_t) = \varphi(R_t, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma})$

In this case, we know exactly how the probability distributions in the output space should look like, where the leftmost output pdf is associated with the rightmost membership function in the input space and vice versa (reversal).

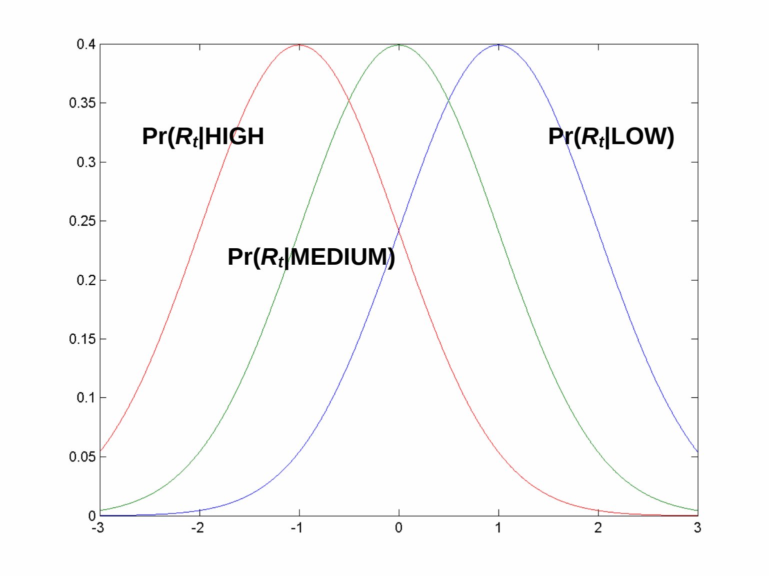


Figure 5.2 - The target output probability distributions for all three rules

The model that is used to estimate these conditional pdf's is a Gaussian mixture model consisting of a single Gaussian. The parameters for the target pdf's would look like this, where **W** is a matrix of weights where each row represents a rule and each column a Gaussian component of the mixture model (one in this case).

W =	1	**M** =	1	**S** =	1
	1		0		1
	1		-1		1

The parameters of this model are estimated using the maximum likelihood criterion as described in paragraph 3.6 and an evolution strategy with population size 100, 50 generations, elitism for 5 individuals, a cross-over probability of 0.7 and a mutation rate of 0.5. The evolution of the maximum likelihood is depicted in figure 5.3.

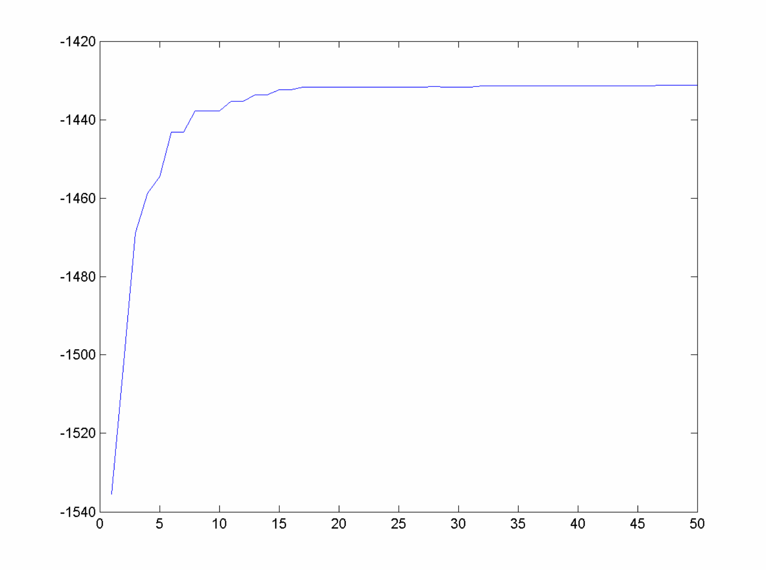


Figure 5.3 - Evolution of the maximum likelihood

The Likelihood does seem to have reached an optimum, according to its asymptotical behavior.

The resulting parameters are:

W =	1	**M** =	0.9370	**S** =	0.9611
	1		0.0682		1.0482
	1		-0.9687		0.9721

Leading to the estimates of the output pdf's depicted in figure 5.4.

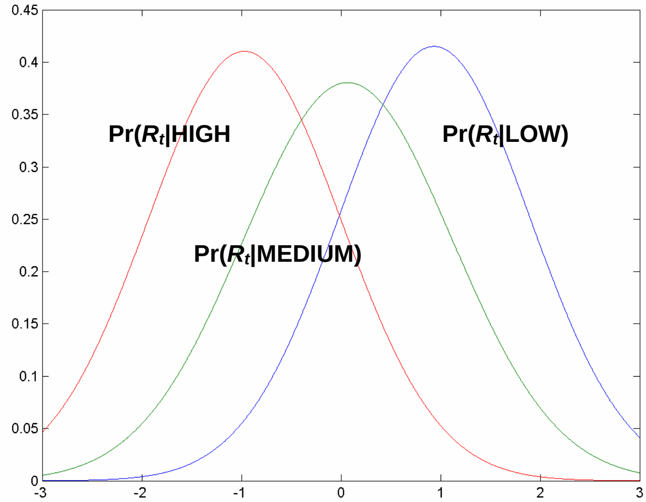


Figure 5.4 - The estimates of the conditional pdf's for the following logreturn R_t

Figure 5.4 shows that the model might have a bias towards smaller standard deviations for smaller samples. The pdf's given "low" and "high" lagged returns are estimated from smaller samples than the one for "medium" lagged returns. This presumption is reinforced by the results from repeating the same experiment.

Interpolations of the pdf's will be made, as described in paragraph 3.7, by interpolating the parameters of the mixture models weighted by the fuzzy memberships of the input lagged returns. For the partitioning of the input space we are currently considering, the membership values for each of the three fuzzy classes of $x = 0.5$ for instance is:

$$\boldsymbol{\mu} = \begin{matrix} 0 \\ 0.5 \\ 0.5 \end{matrix}$$

This means that the interpolated parameters of the target model will be $\mathbf{w} = 1$, $\mathbf{m} = -0.5$, and $\mathbf{s} = 1$.

The interpolated parameters of the model we have found are **w** = 1, **m** = -0.4503, and **s** = 1.0102. With this level of 'reversal' existent in the data, the probabilistic fuzzy approach will obviously outperform a model assuming constant diffusion anytime. Please note that assuming unconditional constant diffusion also leads to an estimation error if we only look at the sample mean and standard deviation. The sample mean in this case is 0.0246, the sample standard deviation is 1.1527, whereas the population mean is 0 and the population standard deviation is 1.

Using a slightly more complex version of the model with more Gaussians, the 'reversal' pattern is also very quickly captured.

5.2 Real-life data

One-minute returns

Looking at the conditional pdf's for one-minute logreturns (Figure 5.5), we can see conditional variances that for the "high" and "low" returns are obviously much larger than for "medium" returns. This is exactly what I predicted in paragraph 2.5 and is certainly not caused by a potential bias towards smaller standard deviations for small samples as discussed in paragraph 5.1.

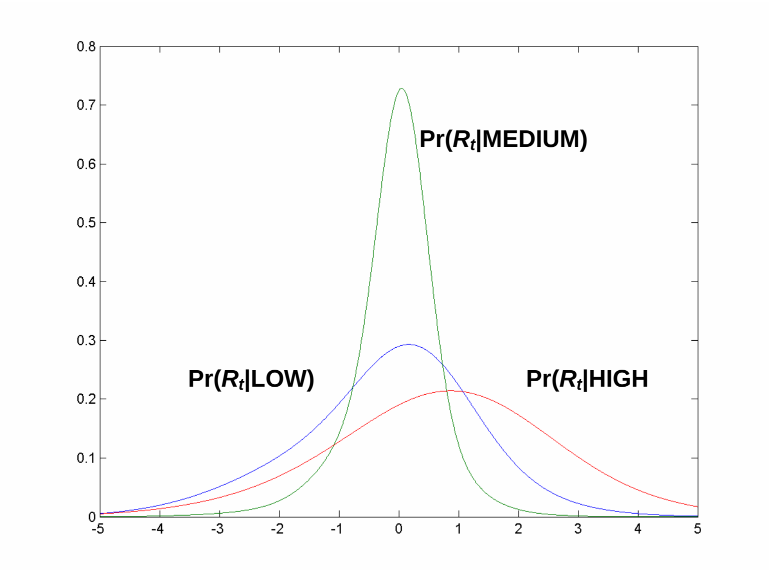


Figure 5.5 – 10,000 one-minute logreturns, 4 Gaussians, after 100 generations

The conditional pdf's for the following logreturn show the resulting means and standard deviations:

$$
\begin{aligned}
&\Pr(R_t \mid \text{"Low"}): && \mu = -0.1767, && \sigma = 1.4428 \\
&\Pr(R_t \mid \text{"Medium"}): && \mu = -0.0524, && \sigma = 0.6513 \\
&\Pr(R_t \mid \text{"High"}): && \mu = 0.7311, && \sigma = 1.9165
\end{aligned}
$$

Contrary to my expectations, the conditional pdf's show a clear tendency to 'drift'. Low returns are more likely to be followed be another low return and even more so, high returns are likely to be followed by high returns. This is exactly the opposite of what I had expected.

Five-minute returns

On five minute returns, the results looked like in figure 5.6. Again the conditional volatilities are exactly like I had expected. However, now there is a conditional probability for 'reversal'. Low or high returns are likely to be corrected by the next return.

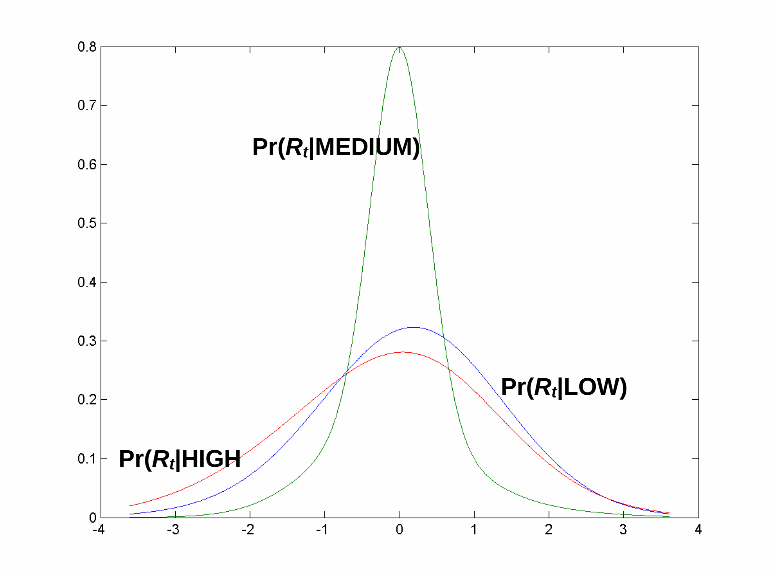


Figure 5.6 – 5000 five-minute logreturns, 3 Gaussians, after 50 generations

The conditional pdf's for the following logreturn show the resulting means and standard deviations:

$$\begin{aligned} &\Pr(R_t \mid \text{"Low"}): && \mu = 0.1292, && \sigma = 1.2404 \\ &\Pr(R_t \mid \text{"Medium"}): && \mu = -0.0084, && \sigma = 0.6138 \\ &\Pr(R_t \mid \text{"High"}): && \mu = -0.1461, && \sigma = 1.3964 \end{aligned}$$

5.3 Simulation results

After simulating 2,750 500-minute stock price paths as described in paragraph 4.4, we can plot a histogram of the resulting logreturns and compare this with the distribution of 500-minute logreturns in a world that adheres to the Black-Scholes assumptions. In this world the 500 minute logreturns are normally distributed with $\mu_{500min} = 100\mu_{5min}$ and $\sigma_{500min} = \sqrt{100}\sigma_{5min}$.

When comparing the simulated stock price paths using the five-minute model with the theoretical distribution of the returns in a Black-Scholes world, it can be seen that the mean of the distribution has shifted a little to the right for the probabilistic fuzzy approach (0.0015 versus -0.0005), whereas the standard deviation of the simulation results is slightly smaller (0.0088 versus 0.0092) and shows some evidence of kurtosis (fat tails).

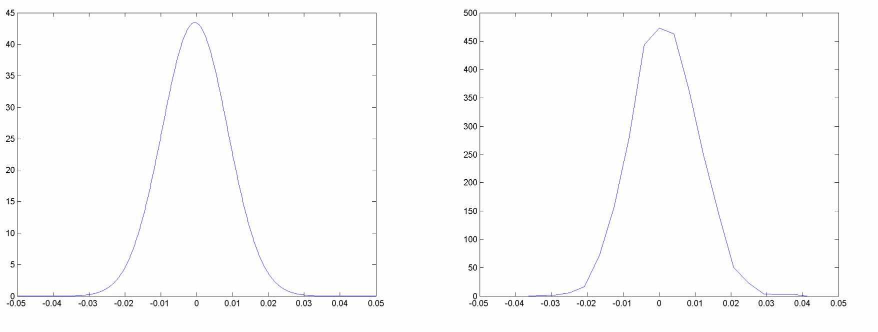


Figure 5.7 - The theoretical distribution (left) versus the simulation results (right)

6. Discussion and further research

6.1 Discussion of the results

The difference in conditional probability distributions found for each of the rules and the way these differences confirm my hypotheses about stock price behavior as explained in paragraph 2.4, shows that a probabilistic fuzzy approach is a meaningful approach to modeling short term stock price behavior. Whether the difference in expected value of the 500-minute return with the Black-Scholes framework is an improvement, needs to be further investigated.
The difference in standard deviation and the 'fat tails' are likely to be an improvement, as the 'fat tails' are a known anomaly of the Black-Scholes model and Corsi et al. (2001) have shown that computing 'realized' returns using high-frequency data leads to structural over-estimation of the longer term volatility.

Looking at the conditional pdf's (Figure 5.6) alone, it is very interesting to see the evidence of a 'reversal' pattern and the increased conditional volatility for extreme returns. These results may lead to a better understanding of the stock pricing process.

However interesting a probabilistic fuzzy approach to modeling stock price behavior, it is one that has serious limitations. The most important of them is the computational expensiveness of both fitting the model parameters and simulating the stock price paths. It does give an option price that is different from the Black-Scholes price, and most likely a better estimation of the expected pay-off. That the model also outperforms the Black-Scholes framework with all the enhancements made over time, including the scaling of realized volatilities (Corsi et al. (2001)) however, is less likely.
The interpretability of the rules however, to my opinion, remains an advantage of the probabilistic fuzzy approach, compared to the adjusted Black-Scholes model.

Looking at the difference between the results for one-minute returns and for five-minute returns with the assumptions from paragraph 2.5 in mind, one may suspect that large shifts in the observed price often take more than a minute to occur. This would explain the tendency to drift that can be seen in the conditional pdf's for one-minute returns.

In the end, one can say that regarding the conditional volatility, the research question in paragraph 1.2 can be answered positively: the difference in conditional volatility for each rule is certainly significant – the standard deviation for "low" and "high" lagged returns is more than twice as large as the one for "medium" lagged returns.
Regarding the mean, this research question is a little bit harder to answer, as conventional statistical theory is not very well suited for this application. Repeating the experiment with a different dataset is the best way to answer this question.

6.2 Further research

There are many possible improvements to the model that is developed in this thesis. I have already mentioned that simulation should probably be done after computing a lookup table. In this paragraph, I will list a few interesting topics for further research.

Exponentially weighting historic returns
In many applications of financial modeling, recent data is assumed to be more representative for future data than less recent data. This concept could easily be implemented in the model used in this thesis by making a small adjustment to the likelihood criterion.

$$L_{adj} = \prod_{t=0}^{T} \Pr\left(R_t \mid R_{t-1}\right) \cdot \lambda^t$$

Where $0 < \lambda < 1$.

American style options
Conditional probabilities for 'reversal' or 'drift' have a large impact on the valuation of American option. In many theories, the probability distributions for

stock returns are always perfectly symmetric. This implies that it is never optimal to execute the option prior to the expiration date. This might change when using a probabilistic fuzzy approach.

Longer lag times or multiple lagged returns

It is very likely that the model in this thesis does not capture all the interdependency within short-term stock returns. To improve this aspect of the model it is possible to either incorporate longer lag times for the input returns, or to use multiple lagged returns as input.

Appendix

Put-call parity

Because there is a direct relation between put and call-option prices, called put-call parity, there is no need to model both types of options. Put prices can simply be deduced from call prices.

Consider two portfolios:

- one European call option plus an amount of cash equal to the strike price discounted at the risk free interest rate ($X e^{-rT}$);
- one European put option plus one share.

Both portfolios are worth max(S_T, X) at expiration. Because they are European, they cannot be exercised prior to expiration. Therefore, the portfolios must have equal value today.

$$c + Xe^{-rT} = p + S_0 \Rightarrow p = c + Xe^{-rT} - S_0$$

The central limit theorem

Many random variables are in practice sums of multiple independent random variables and stock returns are an excellent example. Each daily return is, for instance, the sum of half-hour returns, which are in their turn sums of 30 one-minute returns.

The mean of a sum of random variables is equivalent to the sum of the means of the independent random variables and the variance of the sum is equivalent to the sum of the variances. The central limit theorem states that:

For the sum of a large number of independent random variables from an identical distribution, the sum tends to the normal distribution, whatever the distribution of the summed random variables.

The central limit theorem has a substantial impact on the practice of statistics. Many problems involve sums or averages of random variables, and in these cases, because of this theorem, the normal distribution offers a satisfactory approximation of the true distribution.

One of the implications of the central limit theorem is that, if one assumes stock returns to have a constant volatility and drift and to be independent, that for each sufficiently large interval, the normal distribution is the most appropriate distribution to apply. Another implication is that option prices from binomial trees converge to the Black-Scholes price when the number of steps in the tree grows. Whenever stock returns are not independent or the volatility varies, the central limit theorem is no longer applicable.

Microsoft Corp.

Microsoft Corporation engages in the development, manufacture, license, and support of software products for various computing devices worldwide. Its Client segment offers operating systems for servers, personal computers (PCs), and intelligent devices. The company's Server and Tools segment provides server applications and developer tools, as well as training and certification services.

Its products provide messaging and collaboration, database management, e-commerce, and mobile information access capabilities. It also offers consulting services.

Microsoft's Information Worker segment offers business and personal productivity software applications, including collaboration tools and document management and messaging applications for personal computers.

Its Microsoft Business Solutions segment offers software solutions to manage financial, customer relationship, and supply chain management functions. Its products consist of business solutions, customer relationship management software, retail solutions, and related services.

The company's MSN segment provides online communication and information services, including email and instant messaging, and online search and premium content. It also provides Internet access, and Web and mobile services.
Its Mobile and Embedded Devices segment offers mobile software platform; embedded device software platforms used in consumer electronics devices and enterprise devices; a hosted programmable XML Web service; and software platform to create telematics solutions for vehicles.
Microsoft's Home and Entertainment segment offers the Xbox video game system; PC software games, online games, and console games; television platform products for the interactive television industry; and consumer software and hardware products, such as learning products and services, application software for Macintosh computers, and PC peripherals.
Microsoft was founded in 1975 by William H. Gates III. The company is headquartered in Redmond, Washington.

Key statistics
Market Cap (intraday): 285.92B
Enterprise Value (9-Jan-06)3: 246.39B
Trailing P/E (ttm, intraday): 22.69
Forward P/E (fye 30-Jun-07) 1: 17.67
PEG Ratio (5 yr expected): 1.65
Price/Sales (ttm): 7.10
Price/Book (mrq): 5.93
Enterprise Value/Revenue (ttm)3: 6.11
Enterprise Value/EBITDA (ttm)3: 13.73

Source: www.microsoft.com

References

Andersen and Bollerslev (1998), Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39-4, 885-905.

Bishop, C.M., *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

Black, F. and Scholes, M.S. (1973). The Pricing of options and corporate liablilities. *Journal of Political Economy*, 7, 637 54.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, Vol. 31, 1986, pp. 307-327.

Corsi, F., Zumbach, G., Muller, U. and Dacorogna, M. (2001). Consistent high-precision volatility from highfrequency data, *Economic Notes*, 30, 183-204.

Cox, J., Ross, S., and Rubinstein, M. (1979). Option Pricing: A Simplified Approach. *Journal of Financial Economics*, 7 (October 1979), 229-64.

Derman, E. and Kani, I. (1994). Riding on a smile. *Risk*, vol. 7 no. 2 (February):32-39.

Duffie, D. (1996). *Dynamic Asset Pricing Theory*, 2nd ed. Princeton University Press.

Fama, E.F. (1998) Market efficiency, long-term returns and behavioral finance. *Journal of Financial Economics* 49 (1998) 283Ð306

French, K.R. and Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 17:5–26, 1986.

Hull, J.C. (2003) *Options Futures and other Derivatives*. Prentice Hall.

Jang, J.-S.R, Sun, C.-T., and E. Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*. Prentice Hall.

Kaymak, U., Van den Bergh, W.-M. and Van den Berg, J. (2003) A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. *Proceedings of the 15th belgian-Dutch Conference on Artificial Intelligence (BNAIC'03)*, 187-194, 2003.

Lemieux C, L'Ecuyer P – On the Use of Quasi-Monte Carlo Methods in Computational Finance – *International Conference on Computational Science* (1), 2001

Mamdani (1977). Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, vol. 26, no. 12, pp. 1182-1191, 1977.

Schwefel, H.-P. (1974). Adaptive Mechanismen in der Biologischen Evolution und ihr Einfluss auf der Evolutionsgeschwindigkeit. *Technical Report. Technical University of Berlin. Abschlussbericht zum DFG-Vorhaben* Re 215/2.

Takagi, T. and Sugeno, M. Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116-132, Jan.,/Feb. 1985.

Van den Berg, J., Kaymak, U., and Van den Berg, W.-M. (2004). Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35:291-305, 2004.

Van den Bergh, W.-M., Kaymak, U., and Van den Berg, J. (2002). On the data-driven design of Takagi-Sugeno probabilistic fuzzy systems. *Proceedings of the EUNITE Conference 2002, Portugal*, September 2002

Waltman, L., Kaymak, U., and Van den Berg, J. (2005), Maximum Likelihood parameter estimation in probabilistic fuzzy classifiers.

Waltman, L. (2005), A theoretical analysis of probabilistic fuzzy systems. *Master thesis Informatics & Economics*. Erasmus University Rotterdam, March 2005.

Zadeh, L.A. (1968) Probability measures of fuzzy event. *Journal of Mathematical Analysis and Applications*, 23:421-427, 1968