Some Statistical Properties of Mixture Distribution and Its Applications to Monte Carlo Simulation and Particle Filter

Kangrong Tan*

Abstract

This paper aims to study some statistical properties of mixture distribution, especiallay on its mean, variance, skewness, and kurtosis. From these properties, we find a mixture distribution can provide an accurate approximate for a probability distribution function for observed data, even for a distribution with heavy tails, excess kurtosis, and finite moments. Sometimes, it is important since these phenomena are well observed in financial markets and some scientific fields. Here, in this paper, we also propose an algorithm to generate random numbers from mixture distribution, and it can be utilized in dealing with Monte Carlo simulation and particle filter under the circumstances of mixture distributions. Furthermore, we propose some algorithms to estimate the structure of particle filter and its parameters based upon Genetic Programming and Genetic Algorithm.

1 Introduction

Recent studies have focused on approximating or identifying a probability distribution function for observed data. Some researches suggest Gaussian

* Faculty of Economics, Kurume University, 1635 Miimachi, Kurume City, Fukuoka, Japan

distribution since it can be tackled easily and simply; some researches suggest non-Gaussian stable distribution, or Student t distribution since they have heavy tails [1][2][3]. However, under the circumstances, that a distribution is of excess kurtosis, heavy tails, and finite moments etc, sometimes, it is difficult to fit it by using a single distribution family. In our previous works [4][5], we proposed to fit it by using a mixture distribution. Since a mixture distribution is built by several different distribution families by scaling different parts from their own probability distributions. Thus, it is more flexible to approximate or identify a distribution for observed data than a single distribution family. Sometimes a mixture distribution can avoid serious statistical biases which are introduced by other single distribution family [3][4][5][6][11].

Here, a mixture distribution which we have proposed in our previous researches, is constructed by a weighed sum of a set of radical distribution families, and the distributional parameters of these radical distributions and combinational weights are estimated and optimized by Genetic Algorithm (GA) [4][5][7][8]. As reported in our previous researches [4][5], it is found that an approximate based on a mixture distribution has good statistical properties which are consistent with the real world, and it can also catch the statistical characteristics of the real data. The remainder of this paper is organized as follows. In section 2, we show how to construct a mixture distribution using a set of radical distributions, and discuss some statistical properties of mixture distribution, how to get the parameters optimized by GA, and how to generate random numbers and do other statistical calculations based upon a mixture distribution. In section

3, we show how to do the Monte Carlo simulation under the circumstances of mixture distributions, and its applications to particle filter.

2 Mixture distribution and its properties

2.1 Methodology of mixture distribution

In this section, we propose to approximate or identify a distribution for observed data using mixture distribution. A mixture distribution is a weighed sum of several radical distributions. Here radical ones can be Gaussian, Student t, Generalized Error Distribution (GED), and whatever one thinks it is suitable to the real data. Namely, let

$$r_t \sim \sum_{i=1}^{n} \alpha_i \phi_i \tag{1}$$

and

$$\sum_{i=1}^{n} \alpha_i = 1, (\forall \alpha_i \geq 0) \tag{2}$$

where ϕ_i is one of the radical distributions. A simple case is that r_t is the combination of two normals or one normal and one Student t. Namely, $r_t \sim (1-\alpha)N(\mu_1,\sigma_1^2)+\alpha N(\mu_2,\sigma_2^2)$, or $r_t \sim (1-\alpha)N(\mu_1,\sigma_1^2)+\alpha t(\sigma,\nu)$.

Surely, there are many ways to decide which distribution family should be included in the radical distribution set. Basically, it allows mixture distribution not only catches the characteristics of excess kurtosis much better than simple normals, but also keeps the model to be tackled easily when suitable distributions are chosen as the members of radical distributions. It as well allows mixture distribution has more flexibility to deal with heavy-tailed behavior when Student t, GED etc, heavy-tailed families are chosen as the members of radical distributions.

Such a mixture distribution has some good statistical properties that prevail over those of a single distribution family. Since a mixture distibution has more flexibility to adjust its own distributional shape by scaling different membering distributions by different weights. That is why mixture distribution is much more powerful than single distribution family in coping with observed data.

Plots of p.d.fs of $N(0,1)$, Student with d.f 2.5, Mixture distribution $((1-\alpha)^*N(0,1)+\alpha^*N(0,25))$, Cauchy $\left(\frac{1}{\pi}\frac{1}{1+x^2}\right)$, and N(0,25), where $\alpha = 0.05$, and 0.25 are shown in Figure-1. Seen from the Figure-1, the kurtosis and tail of a mixture distribution can position themselves between N(0,1) and Cauchy distribution, being adjusted by changing the values of α_i and σ_i.

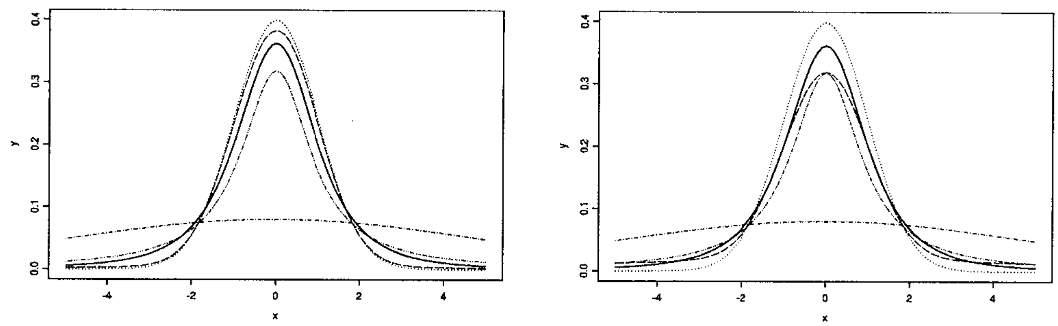

Figure-1: Plots of p.d.fs Left : top to bottom, N(0,1), Mixture, t-2.5, Cauchy, N(0,25). Right : top to bottom, N(0,1), t-2.5, Mixture, Cauchy, N(0,25)

2.2 Properties of mixture distribution

In this section, we study and discuss some statistical properties of mixture distribution. For simplicity and without loss of generalicity, we assume that all the radical distribution functions are continued with the same domains, say, $x \in (-\infty, +\infty)$. Assuming that f can be written in a weighed sum way with n radical distributions as follows,

$$f = \alpha_1 g_1 + \alpha_2 g_2 + \cdots + \alpha_n g_n \tag{3}$$

where $\sum_{i=1}^{n} \alpha_i = 1$, and g_i is a p.d.f. of a radical distribution with mean μ_i and varince σ_i^2.

We discuss some properties of f through the following theorems.

[***Theorem*-1**] Clearly, f is satisfied with the conditions being a p.d.f. Since $f \geq 0$, $\int |f| \leq \infty$, and $\int f = 1$ hold.

[***Theorem*-2**] The mean of f is $\sum_{i=1}^{n} \alpha_i \mu_i$.

[***Proof***] From the definition of mean, it yields,

$$E[x] = \int x f(x) dx \tag{4}$$

$$= \alpha_1 \int x g_1(x) dx + \cdots + \alpha_n \int x g_n(x) dx \tag{5}$$

$$= \alpha_1 \mu_1 + \cdots + \alpha_n \mu_n = \sum_{i=1}^{n} \alpha_i \mu_i \tag{6}$$

where $\mu_1, \ldots, \mu_n$ are the means of $g_1, \ldots, g_n$ respectively.

For simplicity, we just discuss the properties in the case $f = \alpha_1 g_1 + \alpha_2 g_2$, where $\alpha_1 + \alpha_2 = 1$, $n = 2$. It is easy for one to extend them into the case $n = N$ by mathematical induction.

[***Theorem*-3**] The variance of f is $\alpha_1^3 \sigma_1^2 + \alpha_2^3 \sigma_2^2 + \alpha_1 \alpha_2 (\alpha_1 \sigma_2^2 + \alpha_2 \sigma_1^2) + (\mu_1 - \mu_2)^2)$ $+ 2\alpha_1 \alpha_2 (\alpha_1 \sigma_1^2 + \alpha_2 \sigma_2^2)$.

[***Proof***] From the definition of variance, it yields,

$$V[x] = \sigma^2 \tag{7}$$

$$= \int (x - E(x))^2 f(x) dx \tag{8}$$

$$= \alpha_1^3 \sigma_1^2 + \alpha_2^3 \sigma_2^2 + \alpha_1 \alpha_2 (\alpha_1 \sigma_2^2 + \alpha_2 \sigma_1^2 + (\mu_1 - \mu_2)^2) \tag{9}$$

$$+2\alpha_1\alpha_2(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2) \tag{10}$$

[***Remark***] Here if $\mu_1 = \mu_2 = 0$, then $V[x] = \alpha_1\sigma_1^2+\alpha_2\sigma_2^2$.

[***Theorem-4***] The skewness of f is $\frac{\alpha_1\sigma_1^3}{\sigma^3}skewness_1+\frac{\alpha_2\sigma_2^3}{\sigma^3}skewness_2$, when $\mu_{1,2} = 0$

[***Proof***] From the definition of skewness, it yields,

$$skewness = \frac{E(x-\mu)^3}{(\sqrt{V[x]})^3} \tag{11}$$

$$= \frac{\alpha_1 \int (x-(\alpha_1\mu_1+\alpha_2\mu_2))^3 g_1 dx}{\sigma^3} \tag{12}$$

$$+\frac{\alpha_2 \int (x-(\alpha_1\mu_1+(\alpha_2\mu_2))^3 g_2 dx}{\sigma^3} \tag{13}$$

if we assume that $\mu_1 = \mu_2 = 0$ then, it yields,

$$skewness = \frac{\alpha_1 \int x^3 g_1 dx}{\sigma^3}+\frac{\alpha_2 \int x^3 g_2 dx}{\sigma^3} \tag{14}$$

$$= \frac{\alpha_1\sigma_1^3 \int x^3 g_1 dx}{\sigma_1^3\sigma^3}+\frac{\alpha_2\sigma_2^3 \int x^3 g_2 dx}{\sigma_2^3\sigma^3} \tag{15}$$

and since $skewness_1 = \frac{\int x^3 g_1 dx}{\sigma_1^3}$, and $skewness_2 = \frac{\int x^3 g_2 dx}{\sigma_2^3}$, thus,

$$skewness = \frac{\alpha_1\sigma_1^3}{\sigma^3}skewness_1+\frac{\alpha_2\sigma_2^3}{\sigma^3}skewness_2 \tag{16}$$

And it is clear that it is unskewed when both of the $skewness_1$ and $skewness_2$ are equal to zero. On the other hand, it is skewed if $\frac{skewness_1}{skewness_2} \neq -\frac{\alpha_2}{\alpha_1}\frac{\sigma_2^3}{\sigma_1^3}$, $(\forall \alpha_i \neq 0)$ holds.

[***Theorem-5***] The kurtosis of f is $\frac{\alpha_1\sigma_1^4}{\sigma^4}kurtosis_1+\frac{\alpha_2\sigma_2^4}{\sigma^4}kurtosis_2$, when $\mu_{1,2} = 0$

[***Proof***] From the definition of kurtosis, it yields,

$$kurtosis = \frac{E(x-\mu)^4}{(V[x])^2} \tag{17}$$

$$= \frac{\alpha_1 \int (x-(\alpha_1\mu_1+\alpha_2\mu_2))^4 g_1 dx}{\sigma^4} \tag{18}$$

$$+\frac{\alpha_2 \int (x-(\alpha_1\mu_1+\alpha_2\mu_2))^4 g_2 dx}{\sigma^4} \tag{19}$$

since we assume $\mu_1 = \mu_2 = 0$ then, we have

$$kurtosis = \frac{\alpha_1 \int x^4 g_1 dx}{\sigma^4}+\frac{\alpha_2 \int x^4 g_2 dx}{\sigma^4} \tag{20}$$

$$= \frac{\alpha_1\sigma_1^4 \int x^4 g_1 dx}{\sigma_1^4\sigma^4}+\frac{\alpha_2\sigma_2^4 \int x^4 g_2 dx}{\sigma_2^4\sigma^4} \tag{21}$$

and since $kurtosis_1 = \frac{\int x^4 g_1 dx}{\sigma_1^4}$, and $kurtosis_2 = \frac{\int x^4 g_2 dx}{\sigma_2^4}$, thus it yields,

$$kurtosis = \frac{\alpha_1\sigma_1^4}{\sigma^4} kurtosis_1+\frac{\alpha_2\sigma_2^4}{\sigma^4} kurtosis_2 \tag{22}$$

Here we take a look at a special case that is both radical distributions are normals, then $kurtosis_1$, $kurtosis_2 = 3$. And we have the kurtosis of the mixture distribution of two normals, say $kurt_{mix}$, as follows.

$$kurt_{mix} = 3\left(\frac{\alpha_1\sigma_1^4}{\sigma^4}+\frac{\alpha_2\sigma_2^4}{\sigma^4}\right) \tag{23}$$

then the excess kurtosis, say $exkurt_{mix}$, is defined as

$$exkurt_{mix} = 3\left(\frac{\alpha_1\sigma_1^4}{\sigma^4}+\frac{\alpha_2\sigma_2^4}{\sigma^4}\right)-3 \tag{24}$$

$$= 3\left(\frac{\alpha_1\sigma_1^4}{\sigma^4}+\frac{\alpha_2\sigma_2^4}{\sigma^4}-1\right) \tag{25}$$

$$= 3\frac{\alpha_1\sigma_1^4+\alpha_2\sigma_2^4-(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2}{(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2} \tag{26}$$

if excess kurtosis exists then $exkurt_{mix} > 0$, it yields,

$$3\frac{\alpha_1\sigma_1^4+\alpha_2\sigma_2^4-(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2}{(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2} > 0 \tag{27}$$

namely,

$$\frac{\alpha_1\sigma_1^4+\alpha_2\sigma_2^4-(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2}{(\alpha_1\sigma_1^2+\alpha_2\sigma_2^2)^2} > 0 \tag{28}$$

$$\alpha_1\sigma_1^4+\alpha_2\sigma_2^4-\alpha_1^2\sigma_1^4-2\alpha_1\alpha_2\sigma_1^2\sigma_2^2-\alpha_2^2\sigma_2^4 > 0 \tag{29}$$

$$\alpha_1\sigma_1^4(1-\alpha_1)+\alpha_2\sigma_2^4(1-\alpha_2)-2\alpha_1\alpha_2\sigma_1^2\sigma_2^2 > 0 \tag{30}$$

since $\alpha_1+\alpha_2 = 1$, then

$$\alpha_1\alpha_2\sigma_1^4+\alpha_1\alpha_2\sigma_2^4-2\alpha_1\alpha_2\sigma_1^2\sigma_1^2 > 0 \tag{31}$$

namely,

$$\sigma_1^4+\sigma_2^4-2\sigma_1^2\sigma_2^2 > 0 \tag{32}$$

$$(\sigma_1^2-\sigma_2^2)^2 > 0 \tag{33}$$

[***Remark***] It means that the excess kurtosis exists when both of the normals are with different standard deviations, for $\alpha_i \neq 0$. Furthermore, if one of them is a distribution with excess kurtosis, $kurtosis_1 = 3$, $kurtosis_2 > 3$, for example, then it yields,

$$kurtosis = \frac{\alpha_1\sigma_1^4}{\sigma^4}kurtosis_1+\frac{\alpha_2\sigma_2^4}{\sigma^4}kurtosis_2 \tag{34}$$

$$> 3\frac{\alpha_1\sigma_1^4}{\sigma^4}+3\frac{\alpha_2\sigma_2^4}{\sigma^4} \tag{35}$$

then we know the excess kurtosis $exkurt_{mix}$ exists as well.

2.3 Generating random numbers based upon mixture distribution

In the case one wants to calculate a p-value at β percent, such as 1%, 5%, from a mixture distribution, then we propose to use bisection method (binary search algorithm) to do it.

Assuming that the mixture distribution $f(x)$ is given,

$$f(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \cdots + \alpha_n g_n(x) \tag{36}$$

And here what we are going to figure out is a special value x, which satisfies

$$\int_{-\infty}^{x} f(y)\,dy = \beta \tag{37}$$

It yields,

$$\alpha_1 \int_{-\infty}^{x} g_1(y)\,dy + \alpha_2 \int_{-\infty}^{x} g_2(y)\,dy + \cdots + \alpha_n \int_{-\infty}^{x} g_n(y)\,dy = \beta \tag{38}$$

Then we can define $L(x)$ as follows,

$$L(x) = \alpha_1 \int_{-\infty}^{x} g_1(y)\,dy + \alpha_2 \int_{-\infty}^{x} g_2(y)\,dy + \cdots + \alpha_n \int_{-\infty}^{x} g_n(y)\,dy - \beta \tag{39}$$

Before we show our algorithm, we need the following theorem to guarantee our calculation will converge.

[***Theorem***-**6**] $L(x)$ is a continued and monotonic function.

[***Proof***] Assuming that $G_i(x) = \int_{-\infty}^{x} g_i(y)\,dy$, then we have

$$\lim_{x \to x_0} L(x) = \lim_{x \to x_0} \sum_{i=1}^{n} \alpha_i \int_{-\infty}^{x} g_i(y)\,dy - \beta \tag{40}$$

$$\lim_{x \to x_0} L(x) = \lim_{x \to x_0} \sum_{i=1}^{n} \alpha_i G_i(x) - \beta \tag{41}$$

$$\lim_{x \to x_0} L(x) = \sum_{i=1}^{n} \alpha_i \lim_{x \to x_0} G_i(x) - \beta \tag{42}$$

$$\lim_{x \to x_0} L(x) = \sum_{i=1}^{n} \alpha_i G_i(x_0) - \beta \tag{43}$$

$$\lim_{x \to x_0} L(x) = L(x_0) \tag{44}$$

Thus, $L(x)$ is a continued function. On the other hand, assuming that $x_2 > x_1$, then we have

$$L(x_2) - L(x_1) = \sum_{i=1}^{n} \alpha_i \int_{x_1}^{x_2} g_i(y)\,dy \tag{45}$$

since $\forall g_i \geq 0$, $\forall \alpha_i \geq 0$, it yields,

$$L(x_2)-L(x_1) \geq 0 \tag{46}$$

or

$$L(x_2) \geq L(x_1) \tag{47}$$

Thus, $L(x)$ is monotonic function.

So, we can use the following algorithm to do the calculations since $L(x)$ is a continued and monotonic function.

[***Algorithm***-1]

Repeat the following loop till the required accuracy is met.

Step 1) Take x_0, which satisfies $L(x_0) < 0$.

Step 2) Take another x_1, which satisfies $L(x_1) > 0$.

Step 3) Compute $x_2 = \frac{(x_0+x_1)}{2}$ to instead of x_0 or x_1 as the new point according to x_2's value, namely, let $x_0 = x_2$ if $L(x_2) < 0$ and let $x_1 = x_2$ if $L(x_2) > 0$.

Step 4) Evaluate $L(x_2)$ to see if it meets the required accuracy, if it is close enough to zero, then the loop terminates, otherwise it goes back to Step 3 to compute the midpoint again.

A sequence of random numbers from a mixture distribution can be generated by our proposed algorithm while Monte Carlo simulation is required. In this case, what we need to do is to generate a sequence of random numbers β_is based upon uniform distribution, and then do the above-proposed algorithm to get a sequence of random numbers x_is, which are distributed by the mixture distribution. We show the applications in section 3.

2.4 Estimation of parameters and weights

The parameters in a mixture distribution should be estimated and optimized. There serval ways to do it. Here, we propose to use GA to realize this goal, since GA is one of the most powerful optimization methods, which usually converges in a global optimal solution, not like the others which sometimes converge in a local one [5][7][8].

The basics of GA are summarized as follows.

[***Algorithm***-2]

Step 1) Generate individuals by random numbers as the first generation with a certain population, each individual represents a group of distributional parameters and combinational weights.

Step 2) Evaluate each individual by predetermined fitness function to get its fitness value.

Step 3) Select two individuals with higher fitness values from the present generation, and to apply genetic operations (crossover, mutation operation) to them to reproduce new individuals (offsprings) as the next generation.

Step 4) Again evaluate the fitness of each individual of the new generation.

Step 5) See if the results meet the *Termination Conditions* (i.e., repeating times, or error range), if so, then GA terminates, otherwise it goes back to Step 3.

Usually there are two ways to do it. One is to use the likelihood function to define the fitness function. Namely,

$$L(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{48}$$

Thus, the fitness function for jth indiviual can be defined as

$$Fitness_j = \frac{L(x;\theta)}{\sum_j L(x;\theta)} \tag{49}$$

where f is p.d.f of a mixture distribution, θ and x_i are a set of parameters and an observation respectively.

The other way is to use the square root error to define the fitness function. Namely,

$$Fitness_j = \frac{\epsilon(x;\theta)}{\sum_j \epsilon(x;\theta)} \tag{50}$$

where $\epsilon = \Sigma(\hat{f}-f)^2$. Where $\hat{f}$, and f are the empirical distribution and estimated distribution respectively. Example are shown in our previous works [4][5].

3 Mixture distribution based Monte Carlo simulation

In this section, we present some applications and their results of Monte Carlo simulation based on mixture distributions. In section 3.1, we show how to evaluate the VaR if one confronts with some mixture distributions. In section 3.2, we show how mixture distribution can be applied to particle filter.

3.1 Evaluating the VaR

Firstly, supposing that we have several different mixture distributions shown in Table-1. We therefore generate a lot of random numbers based upon these mixture distributions according to the ***Algorithm*-1** that we have shown in 2.3, and then to evaluate the VaR by those random numbers. The VaR is estimated as follows [9][10][11].

$$R_t = \mu + \sigma Z_i \tag{51}$$

$$VaR = -R_k^* \tag{52}$$

$$k = int(\alpha * n) \tag{53}$$

where $R^* = \{R_1^*, \cdots, R_n^*\}$, which is sorted in ascending order.

	Type
Mixture-1	$0.5*N(0,1)+0.5*N(0.5,1)$
Mixture-2	$0.8*N(0,1)+0.2*N(0.5,1)$
Mixture-3	$0.3*N(0,2^2)+0.2*N(0,1^2)+0.5*N(0.5,1.5^2)$

Table-1: Types of mixture distributions

The size of the random numbers we generate is 10,000. And the simulation results of the VaR are shown in the following Table-2.

VaR	1 %	5 %
Mixture-1	2.1798	1.4582
Mixture-2	2.2617	1.5641
Mixture-3	3.8369	2.4416

Table-2: Results of simulation of the VaR

3.2 Particle filter with noise of mixture distribution

Recently particle filter has been studied for various purposes and fields, such as, tracing a moving object, figuring out some parameters in a dynamic system [12][13][14]. It can be described as follows.

$$x_t = f(x_{t-1}) + w_t \quad (54)$$

$$y_t = g(x_t) + v_t \quad (55)$$

It is known as Kalman Filter when w_t, and v_t are Gaussian. However, in the real world, the conditions which both w_t and v_t are Gaussian are rare, in many cases they are not. Thus, particle filter has been suggested in dealing with those dynamic systems with non-Gaussian noises. It has been widely applied from data engineering to social problems.

In this section, we discuss two types of its applications. One is the noise distributions are given, namely, w_t and v_t are known, but the structure of filter, $f(x)$ and $g(x)$ are unknown. In this case, we can estimate the filter's structure by using Genetic Programming (GP), for simplicity, we call it TYPE-I problem. Another one is the noise distributions are unknown, but the structure of the filter $f(x)$ and $g(x)$ are given. In this case, we can estimate the noise distributions by using GA, for simplicity, we call it TYPE-II problem. We show how to deal with these two different problems in the following subsections.

3.2.1 Type-I problem

In this section, we show how to estimate the structure of a particle filter by using GP. Assuming that x_0 is known, for simplicity, we can use the following algorithm to estimate the structure of the filter $f(x)$ and $g(x)$. However, w_t and v_t are given.

[***Algorithm***-3]

Step 1) to draw N samples from a mixture distribution w_t, say, $w_{1|0,i}$, for i=1...N. Then to generate N values of $x_{1|0,i}$ by using $w_{1|0,i}$, based upon

$$x_t = f(x_{t-1}) + w_t \tag{56}$$

where $f(x)$ is provided by a GP's individual.

Step 2) since $y^T = \{y_t\}_{t=1}^T$ is given, then y_1 and $x_{1|0,i}$ can be used to get N values of $v_{1|0,i}$, namely,

$$v_{1|0,i} = y_1 - g(x_{1|0,i}) \tag{57}$$

where g(x) is provided by the same GP's individual.

Step 3) to calculate the relative probability of every $v_{1|0,i}$ by using $v_1 \sim v_t$, where v_t is given. Namely,

$$q_1^i = \frac{p(v_{1|0,i})}{\Sigma p(v_{1|0,i})} \tag{58}$$

Step 4) to resample N values with replacement of $x_{1|0,i}$ with relative weight q_1^i, say, x_1^i.

Step 5) to go to step 1. To draw new samples from w_t, say, $w_{2|1,i}$, as it is shown in step 1), then $x_{2|1,i}$ is obtained. And to repeat step 2 to step 4 as well, till the end of the sample y_t.

Thus, we get the likelihood function as follows.

$$p(y^T) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(v_{t|t-1,i}) \tag{59}$$

From the likelihood function, we can see which individual of GP's generation provides the better estimates. We define the fitness function just like in 2.4 section, to schedule the next generation. So, we can get the optimal estimates of $f(x)$, and $g(x)$ from GP, when GP terminates.

3.2.2 Type-II problem

In this section, we show how to use the GA algorithm to estimate the parameters and discuss some applications. Assuming that x_0 is known, for simplicity, we can use the following algorithm to estimate the mixture distribution.

[*Algorithm*-4]

Step 1) to draw N samples from a mixture distribution w_t provided by GA's an individual. Say, $w_{1|0,i}$, for i=1...N. Then to generate N values of $x_{1|0,i}$ by using $w_{1|0,i}$, based upon

$$x_t = f(x_{t-1}) + w_t \tag{60}$$

Step 2) since $y^T = \{y_t\}_{t=1}^T$ is given, then y_1 and $x_{1|0,i}$ can be used to get N values of $v_{1|0,i}$, namely,

$$v_{1|0,i} = y_1 - g(x_{1|0,i}) \tag{61}$$

Step 3) to calculate the relative probability of every $v_{1|0,i}$ by using $v_1 \sim v_t$, where v_t is provided by GA's an individual. Namely,

$$q_1^i = \frac{p(v_{1|0,i})}{\Sigma p(v_{1|0,i})} \tag{62}$$

Step 4) to resample N values with replacement of $x_{1|0,i}$ with relative weight q_1^i, say, x_1^i.

Step 5) to go to step 1. To draw new samples from w_t, say, $w_{2|1,i}$, as it is shown in step 1), then $x_{2|1,i}$ is obtained. And to repeat step 2 to step 4 as well, till the end of the sample y_t.

Thus, we get the likelihood function as follows.

$$p(y^T) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(v_{t|t-1,i}) \tag{63}$$

From the likelihood function, we can see which individual of GA's generation provides the better estimates. We define the fitness function just like in 2.4 section, to schedule the next generation. So, we can get the optimal estimates of w_t, and v_t from GA, when GA terminates.

Here is an application of particle filter, which has mixture distribution $w_t \sim 0.2{*}N(0,1) + 0.8{*}N(0,2)$, and $v_t \sim N(0,1)$. Seen from Figure-2, it gives an inference track through the whole time span under the circumstances of mixture noise.

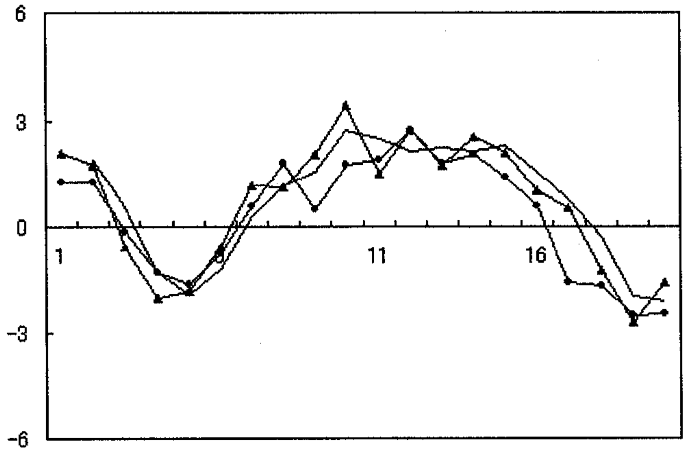

Figure-2:
triangle points: original data
dot points: estimated data
smooth points: smoothed data

4 Conclusions

In this paper, we have studied and discussed some properties of mixture distribution. We have found that it is more powerful to use a mixture distribution to approximate or identify a probability distribution from observed data sets than a single probability distribution family. By introducing radical distributions with heavy tails, a mixture distibution shows better performance than single distribution family in catching the heavy-tailed behavior and excess kurtosis. And we have shown how to get the parameters in a mixture distribution estimated by GA or GP. In our numerical applications, we have shown the applications of mixture distributions in evaluating the VaR, and the application of particle filter with mixture noise. And we have proposed to estimate Type-I, and Type-II problems using GP and GA respectively. We have found that mixture distribution is a good tool to approximate or identify a distribution, since it has the flexibility to adjust its shape to be more suitable for the shape of an empirical distribution of observed data sets by scaling different parts of radical distributions by different weights.

5 Acknowledgements

The author would like to thank the organizing committee members of Journal of the Society for Studies on Industrial Economies for their kind suggestions and arrangements. This research is partly supported by the Japan Society for the Promption of Science under the grant number (B) 19310099 and (C) 19510164, and the author would like to thank the organization.

References

[1] Campbell, J. Y., Lo, A. W., The econometrics of financial markets, Princeton Press, 1997.

[2] Fama, E. F., "Portfolio analysis in a stable Paretian market", *Management Science*, vol.11, pp.404-419, 1965.

[3] Tan, K., "Fractality in financial market", *Journal of the society for studies on industrial economies*, vol.43, 3, pp.553-566, 2002b.

[4] Tan, K., "Modeling Returns Distribution Based on Radical Normal Distributions", *Journal of the society for studies on industrial economies*, vol.46, 3, pp.449-467, 2005.

[5] Tan, K., Tokinaga, S., "Identifing returns distribution by using mixture distribution optimized by genetic algorithm", *Proceeding of NOLTA*2006, 119-122, 2006.

[6] Hull, J., Options, futures, & other derivatives, Prentice Hall, 1997.

[7] Goldberg, D., Genetic algorithm: in search, optimization, and machine learning, Addison-Wesley Press, 1989.

[8] Tan K., Tokinaga S., "Optimization of fuzzy inference rules by using the genetic algorithm and its application to bond rating", *Journal of Operations Research Japan*, vol.42, 3, pp.302-315. 1999.

[9] Jorion, P., 1996. Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill.

[10] Tan, K., "Evaluation of VaR in the Japanese financial market", *Journal of the society for studies on industrial economies*, vol.43, 2, pp.449-467, 2002a.

[11] Tsay, R. S., Analysis of financial time series, John Wiley & Sons, 2002.

[12] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing, 50(2):174. 188, February 2002.

[13] A. Doucet, S. J. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing, 10:197.208, 2000.

[14] M. Isard and A. Blake. Condensation . conditional density propagation for visual tracking. Int. Journal of Computer Vision, 29(1):5.28, 1998.