Approximating Probability Distribution Function based upon Mixture Distribution Optimized by Genetic Algorithm and its Application to Tail Distribution Analysis using Importance Sampling Method

Tan, Kangrong
Faculty of Economics, Kurume University

Tokinaga, Shozo
Department of Economic Engineering, Faculty of Economics, Kyushu University : Professor

https://doi.org/10.15017/15741

出版情報：經濟學研究. 74 (1), pp.183-196, 2007-09-28. 九州大学経済学会

Approximating Probability Distribution Function based upon Mixture Distribution Optimized by Genetic Algorithm and its Application to Tail Distribution Analysis using Importance Sampling Method

Kangrong Tan　　Shozo Tokinaga

1 Introduction

The problem to approximate a probability distribution function (p.d.f) is still a basic and crucial task in various fields, such as, engineering, economics and finance as well as statistics [1][2][3][4][5][9][10][13][14][17][19]. Especially for the probability analysis corresponding to rare events, it is necessary to focus on the tail behavior of the distribution function, and sometimes we are puzzled by so-called fat-tailed distributions [1][5][3][9][12]. Fat-tailed behavior leads to overestimation or underestimation of rare events, and serious accidents, such as, large packet losses in network traffic, and large asset losses in financial markets. So, an accurate approximation method for p.d.f is necessitated.

This paper deals with an approximation method based upon mixture distribution, which is optimized by Genetic Algorithm (GA) and its application to tail distribution analysis by using the IS (Importance Sampling) [14][15][16][17]. For an approximate of p.d.f, due to good statistical properties and tractabilities, conventional distributions, such as normal, lognormal, and non-Gaussian stable distribution and their extensions are well used in theoretical research and numerical approaches. However, as shown from many empirical results, assumptions for these distributions are inconsistent with the phenomena, since many distributions are observed with excess kurtoses, fat tails, skewnesses, and finite moments [5].

In conventional works, we find several successful applications using probability distribution function to approximate p.d.f, which are optimized by the GA [2][4][19][13][17]. One of them is to optimize two typical distributions (Gamma and log-normal distributions) by the GA, as to get the approximations of p.d.fs for natural water flow under different situations. But, in this case, the complexity of the distribution is limited, and mixture distribution is not considered by the authors [2]. One of them is to predict the error distribution or to generate random numbers, the combination of multiple p.d.fs is utilized and the weights are optimized by the GA [4][19]. But, the mixture distribution is oriented only for the error estimation and random number generation, so the radical p.d.fs are limited. It is also reported in the pattern recognition that the GA method for searching the optimal parameters based on the nonparametric estimation in mixture distribution is better than the conventional method, such as Expectation Maximization (EM) [13].

Thus as to generate relevant functional form, we propose to approximate a distribution by a mixed distribution function optimized by the GA[14][15][16][17][18]. Here, mixture distribution is referred to a weighed sum of a set of radical distributions, and whose distributional parameters of the radical distributions and combinational weights are optimized by the GA. Then, as to examine the capability of our proposed method, estimation problem of returns distribution of stock prices is applied and discussed later in this paper. The results are compared with the results obtained by conventional methods, such as normal distributions. Moreover, tail distribution (probability of rare event) which is an important task in asset management is evaluated based on the method of mixture distributions, and its efficiency and accuracy can be improved by using the IS method [6][7][16][18]. As a result, we find that our proposed approximation method provides an accurate approximate for p.d.f and its application in the risk analysis in asset allocation.

The rest of this paper is orgnized as follows. Section 2 shows the basic idea for approximating p.d.f by using several radical distributions, whose parameters are optimized by the GA. Section 3 gives the overview of IS method for improving the efficiency and accuracy of estimation of tail distribution. Section 4 shows some applications and their numerical results with real market data sets.

2 Approximation of p.d.f using GA

2.1 Why mixture distributions

At first, we concisely summarize the limitations of the conventional distributions, normal, lognormal, and non-Gaussian stable. Normal distribution has a tractable property, but it is claimed that there no low bound -1 exists in normal distribution. Furthermore, it is not supported by many empirical results since normals can not catch excess kurtosis, fat tails. The lognormal distribution has the same limitation, though, it solves the low bound problem. Another kind of p.d.f like non-Gaussian stable distribution allows the sum of probabilistic variables still be a stable distribution. And it can catch excess kurtosis and fat-tailed behavior in many cases. But, the problem is that non-Gaussian stable distribution has infinite moments. The estimates of variance and kurtosis tend to be larger and larger and not to converge even though the sample size increases. It doesn't match the reality of practice when finite moments are observed. Furthermore, it is complicated when applied to risk measurement and risk management.

There are many variates of these conventional p.d.fs and their combinations, but the main problem is how to optimize the parameters included in the p.d.f approximation [5]. Therefore, it is necessary to consider more general optimal approaches for the functional approximation.

Several approaches are usually utilized to optimize parameters included in the evaluation function, such as, the Steepest Descent Algorithm for monotonic functions, Back Propagation Algorithm for finding weights in neural networks. However, the GA is considered to be one of the most powerful and robust tools for optimizing the evaluation function, even in the case of nonlinear evaluation functions, or multi-peak functions [8][14][15][17]. Thus, the GA is chosen here to optimize the approximation function.

We must at first explain the following items before we show the applications, namely, the types of radical (basic) functions, the string in the GA individuals, and the approximation evaluation.

(1) mixture distribution of radical distributions

We propose to approximate p.d.f as a weighed sum of radical distributions (mixture distribution). Especially we choose several different families of p.d.fs as the radical functions, such as, normal distributions with different means and variances, and Student ts with different degrees of freedom. Thus, a p.d.f can be represented using the variable $r(x)$ as

$$r(x) \sim \sum_{i=1}^{n} \beta_i \phi_i(x) \tag{1}$$

where the constraint is

$$\sum_{i=1}^{n} \beta_i = 1 \tag{2}$$

where ϕ_i is a radical distribution, a chosen member from a p.d.f family.

It is expected that mixture distribution can represent the characteristics of excess kurtosis and fat tails much better than the conventional normal distribution, and it also has several good statistical properties, such as finite moments when the chosen radical ones have finite moments.

(2) GA string

As mentioned from previous discussion, the variables necessary for the approximation of a p.d.f are the parameters included in each radical function and the weights in mixture distribution. These variables are embedded in the string of individuals in the GA.

For simplicity, we denote the parameters of the radical functions as $\phi_1, \phi_2, ..., \phi_K$. The weights among the radical functions are denoted as $w_1, w_2, ..., w_K$. The individuals are composed of these variables whose lengths are identical.

Moreover, as to obtain approximation at an arbitrarily attainable level, we propose to design the GA system with individuals as length-enlargable ones. If a problem can not be solved within a predetermined accuracy using the current lengths of the individuals, the length (bits) of individuals are automatically to be enlarged, so that it turns to be possible to represent more radical distributions with an improved accuracy of approximation.

(3) Fitness

We must define the fitness of an individual in the pool used for the GA. It is based that the functional values of discrete points (denoted as f_i) of the variables for p.d.f are available for the approximation. So, we have a set of estimated (calculated) values $\hat{f}_i$ for approximation which are obtained form the approximated mixture distribution. Then, we have the root mean square error e_{ij} between f_i and $\hat{f}_i$ for the case where we obtain $\hat{f}_i$ by using the j th individual in the pool. Finally, we define the fitness function for evaluating jth individual as follows.

$$Fitness_j = \frac{1}{(n-1)}\left(1 - \frac{\sum_i e_{ij}^2}{\sum_j \sum_i e_{ij}^2}\right) \tag{3}$$

2.2 Optimizing by the GA

GA is known as one of the efficient optimization methods, which converges hardly in a *local* optimal solution while searching for a *global* optimal one. It has been widely applied in many problems ranging from scientific studies to social studies [8].

The basics of our GA are described as follows.

Step 1: Generation of initial individuals

At first, we generate individuals by random numbers as the first generation. Each individual represents a set of parameters of radical functions and combinational weights.

Step 2: Evaluation of fitness

Then, we evaluate each individual by predetermined fitness function, and sort them according to the values of their fitness.

Step 3: Genetic operations

We select two individuals with higher fitness values from the present generation at a certain predetermined probability. The selection strategy has a large of variations, one of well-used strategies is roulette strategy. And then to apply genetic operations (crossover, mutation operation) to them to reproduce new individuals (offsprings) as the next generation. A crossover operation is referred to randomly decide crossover positions on the two selected individuals, then to exchange parts of two individuals each other. A mutation operation is referred to randomly decide mutation positions under a certain probability, and change those position values of an individual.

Step 4: Termination

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

We show an example of p.d.f approximation by using the GA. We assume the p.d.f is known and is given as a mixture function $y(x) = 0.49 * N(-2, 1) + 0.21 * N(2, 1) + 0.3 * Studentt(d.f = 3)$. Seen from Fig.1, the p.d.f is characterized by the shape having two tops with fat tails, and it is difficult to be approximated by a combination of ordinary normal distributions. But, we have the final result of approximation with running of 2000 generations in GA, the estimated form is $y(x) = 0.50 * N(-1.99, 1.01) + 0.21 * N(1.98, 1.07) + 0.298 * Studentt(d.f = 3.18)$ and the root mean square error of the approximation is 0.00049349. The result shows an example of its acurrate approximation for p.d.f.

2.3 Analysis of tail distribution

In the previous sections, we simply propose an approximation method for the p.d.f based on mixture distribution which is optimized by the GA. We find there is another meaningful task in the analysis of tail distribution when we get the explicit expression of the mixture distribution.

For example, in the network traffic, link delay depends on the extreme values of delay distributions rather than the whole functional forms, which affects direct the packet losses in networks [3][9]. Similarly, almost all financial institutions are suggested by the authorities to evaluate the risk of asset losses in the next period, such as one week [10][12][6][7]. The quantile α of the loss expectation is defined as the Value at Risk (VaR), and usually the value of α is set to between 1%

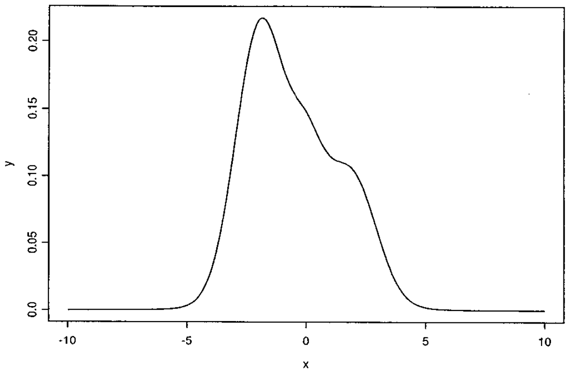


Figure 1: Mixture distribution constructed with two normals and a Student t

to 5%. Here, we use the approximate of mixture distribution as the real p.d.f for the evaluation of tail distribution [8][14][15][17].

3 Improved tail estimation by IS

3.1 Basics of IS

For evaluation of tail distribution of an approximated p.d.f, there exists several methods, such as the direct integration of each radical distributional functions composing the p.d.f, and the Monte Carlo simulation as an alternative. In this paper, we use the Monte Carlo method but make some extension by using the IS to improve the estimation results [8][9][11][16][18].

The IS method exploits tail distribution effectively using the transformation of the distribution function, so that the extreme shape (called tail) of the distribution can be figured out more accurately. The IS method is widely adopted to estimate parameters in systems such as the delay distribution in networks as well as the fractal time series and surfaces [11][18]. It is confirmed by many simulation studies, that the IS method remarkably improves the efficiency of estimation of tail distribution compared to conventional Monte Carlo simulation.

For a given probability p, the tail distribution x_p is defined to be the $(1-p)$th quantile of the distribution.

$$P(x > x_p) = p \tag{4}$$

For simplicity, we assume that the probability $P(x)$ is the distribution of loss in asset allocation during a certain time period, and x_p is a large number in case of estimating $P(x > x_p)$.

The computational cost required to obtain accurate Monte Carlo estimates of tail distribution is often enormous, since a large number of runs (asset loss evaluations, for example) are required to obtain accurate estimates of the loss from the distribution in the region of interest. Thus we apply the IS by changing the measure for sampling probability(the change of loss distribution).

The IS is an appropriate technique for rare event simulations. The standard simulation is inaccurate for estimating $P(x > x_p)$, while for large x_p there few samples are obtained in the region of interest, where $x \simeq x_p$. Effective IS generates a disproportionally large number of samples in the region of interest. For the IS, we have $P(x > x_p) = E[I(z > x_p)l(z)$ where $E(.)$ means the expectation under the IS distribution, and $l(z)$ is the likelihood ratio with new random variable z.

3.2 Selecting Importance function

Under asset loss estimation, we denote

$$p_t = P(x > x_p) = \int_{x_p}^{\infty} f(x)dx \tag{5}$$

where the function $f(x)$ is the p.d.f of asset loss x in a certain period. Then, we introduce a distribution function to generate twisted random variable z as follows.

$$p_t = \int_{x_p}^{\infty} f^*(z)\frac{f(z)}{f^*(z)}dz \tag{6}$$

where the function $f^*(x)$ is the probability density function (called Importance Function) for the twisted variable z having higher probability in the region close to x_p and greater than x_p.

Then, we define the Importance Function $f^*(x)$ which is used as an alternative density function or a biasing density function. Biasing by exponential twisting is most easily explained by means of derivation of statistical upper bounds on tail probability, and is widely used various areas such as information theory.

The basics of the exponential twisting is summarized as follows.

(1) function $f^*(x)$

The function $f^*(x)$ is defined as

$$f^*(x) = e^{sx-\mu(s)}f(x) \tag{7}$$

where the function $\mu(s) = \log M(s)$ is obtained from the moment generating function $M(s) = E\{\exp(sX)\}$ of $f(x)$.

(2) optimal value of s

If the functional form of $f(x)$ is given, then the optimal value of s is determined so that the two quantities are to be minimized.

$$I(s) = \int_{t}^{\infty} e^{-(sx-\mu(s))}f(x)dx \tag{8}$$

$$\bar{I}(s) = e^{-2(st-\mu(s))} \tag{9}$$

If we choose the function $f(x)$ an exponential density function $f(x) = \alpha e^{\alpha x}$, then we have optimal value of s as s_t

$$s_t = \alpha - \frac{1}{t} \tag{10}$$

For example, if we choose as $\alpha = 1$ and $p_t = 10^{-6}$, then we have $t = 13.81551$.

(3) approximation of function

We assume that the p.d.f is approximated by a function $g(x)$ for the variable x based on the mixture distribution. Then, we can get the probability by substituting the value of x into the function $g(x)$.

(4) giving the gain of IS

We also denote the gain of IS as follows.

$$\tau = \frac{p_t(1 - p_t)}{E^*\{1^2(X \geq t)l^2(X)\} - p_t^2} \tag{11}$$

It means the ratio of the sample sizes of standard Monte Carlo method and importance sampling method under the circumstance with the same estimator variances.

We can recognize the ability of the IS by a simple example. For a given exponential distribution $f(x)$ with $\alpha = 3$, we generate $M = 500000$ samples along the distribution. Then, we calculate the probability of rare event $p_t = P(x > x_p) = 10^{-6}$. We have the estimation for expected value of p_t ($E\{p_t\}$) and its variances ($V\{p_t\}$) as ($E\{p_t\} = 0.90001E^{-06}$, $V\{p_t\} = 1.46316E^{-12}$. On the other hand, we obtain estimation result by using the IS ($E\{p_t\} = 1.00180E^{-06}$, $V\{p_t\} = 9.92378E^{-17}$) with the sample size $M = 100000$. At the same time, we have the value τ for the example as $\tau = 54075.94$. The fact shows the ability of IS to increase the estimation result.

4 Applications

4.1 Stock returns

In this section, we present some numerical applications for the approximation of p.d.f. we estimate the mixture distributions for the stock prices in real markets, where one p.d.f is almost symmetric, and the other is asymmetric. Fig.2 shows the time series of these stock prices.

We prepare two data sets, called Data A and Data B from these stock prices. The Data A is composed of the logarithm of relative returns of daily stock price of Tosho (in Fig.2 (left)), from April 1, 1983 ~ September 21, 2001. The Data B is composed of the logarithm of relative returns of daily stock price of IBM (in Fig.2 (right)), from April 1, 1983 ~ September 21, 2001. The descriptive statistics for the returns of Data A and Data B is summarized in Table 1. The Skewnesses of A and B are -0.34628 and -7.95617 respectively, A is almost symmetric, B is asymmetric. And the kurtoses of A and B are 12.66296, 128.1785 respectively, excess kurtoses are observed here.

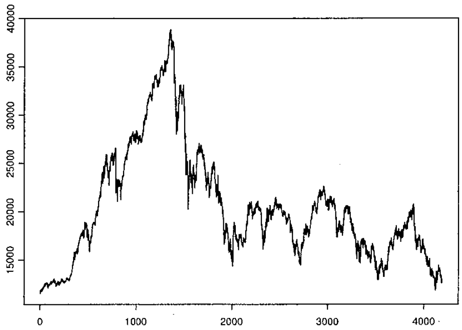
The conditions of the GA procedure are given as follows.

Number of individuals:100

Crossover probability:0.27

Mutation probability:0.15

The ranges of parameters included in the radical functions are given as. $\sigma_i \in (\sigma_{min}, \sigma_{max})$ as (0.1, 50) and $\mu_i \in (\mu_{min}, \mu_{max})$ as (-5, 5).



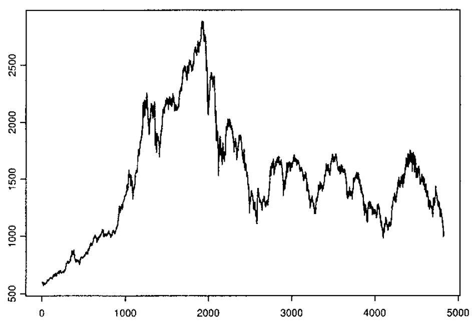


Figure 2: Prices for Data A and B

Table 1. Statistics of returns for Data A and Data B

	Data A	Data B
Mean	4.693E-05	-6.4E-04
Median	7.107E-05	6.06E-04
S.D	0.0050128	0.031477
Kurtosis	12.662960	128.1785
Skewness	-0.346276	-7.95617
Range	0.1082526	0.739272
Min	-0.068663	-0.60809
Max	0.0395896	0.131184
Sum	0.2264266	-6.91786
Size	4825	10800

The first mixture distribution is estimated by GA as a weighed sum of three normal distributions. The estimated parameters are listed in Table 2. Namely, $r_t \sim \beta_i \Sigma_{i=1}^{3} N(\mu_i, \sigma_i^2)$. And the estimated error is 0.00011793, which is the sum of the approximation error $\sum e_{ij}$ obtained by the individual of the highest fitness at the final generation of the GA.

Table 2. Estimation results for Data A

	$N(\mu_1,\sigma_1^2)$	$N(\mu_2,\sigma_2^2)$	$N(\mu_3,\sigma_3^2)$
β_i	0.916708410	0.072212338	0.010278450
μ_i	0.002639055	-0.30707550	-0.21727777
σ_i	1.090306759	2.594726324	3.292275429

The second mixture distribution is estimated by GA as a weighed sum of three normal distributions plus a Student t. The estimated parameters are listed in Table 3. Namely, $r_t \sim \beta_i \Sigma_{i=1}^{3} N(\mu_i, \sigma_i^2) + t(\nu)$. And the estimated error is 1.12762E-05, which is the sum of the approximation error $\sum e_{ij}$ obtained by the individual of the highest fitness at the final generation of the GA.

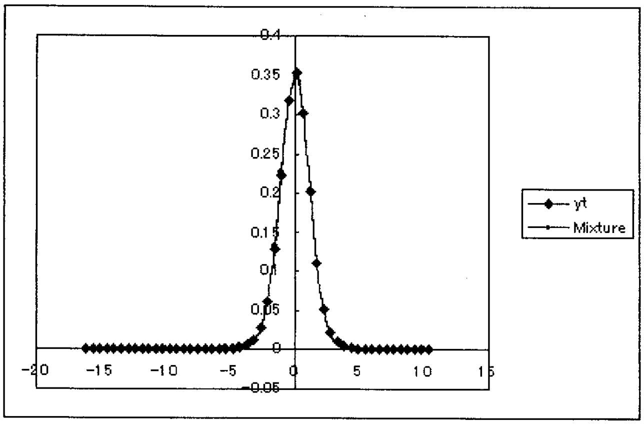


Figure 3: Plots of p.d.fs; solidline: estimated mixture distributions; dashline: standard normal distributions; points: observations.

Table 3. Estimation results for Data B

	$N(\mu_1,\sigma_1^2)$	$N(\mu_2,\sigma_2^2)$	$N(\mu_3,\sigma_3^2)$	$t(\nu)$
β_i	0.183533490	0.009949307	0.024255246	0.782261968
μ_i	0.224407910	0.339988232	0.262051344	0
σ_i	1.000061512	4.187711720	1.010577679	1.090184392
ν	-	-	-	24.17678642

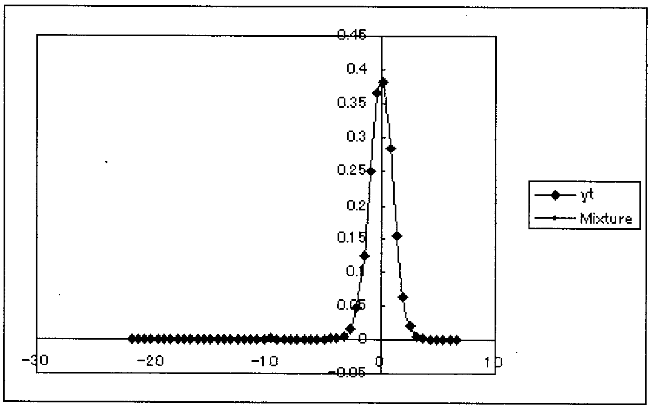


Figure 4: Plots of p.d.fs; solidline: estimated mixture distributions; dashline: standard normal distributions; points: observations.

Fig.3 and fig.4 show the estimated mixture distributions obtained by the GA. In the figures, the solid-lines mean the estimated mixture distributions, points mean the distributions of observations. As seen from Fig.3, 4 clearly, mixture distributions are much more closer to observations than the conventional normal distributions. Mixture distributions are seemed to catch excess kurtosis and fat tails behavior much better than normal distributions. Additionally, the fitness and prediction errors have converged after GAs have been implemented 50 generations. Fig.5 shows the changes of fitness and error in last generation.

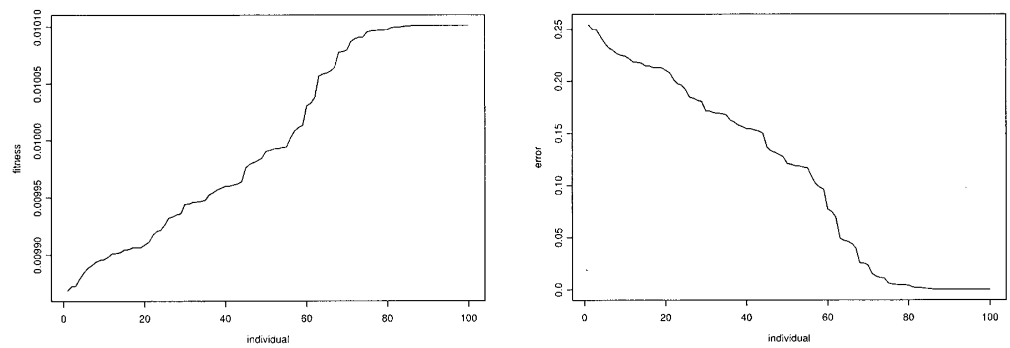


Figure 5: Plots of fitnesse and errors of last generation

4.2 Tail estimation using IS

In this section, we examine the estimation of tail distribution (percentile of loss probability) obtained by the IS method proposed in this paper by comparing with the results obtained solely from the Monte Carlo method [6][7][8][14][15][16][17].

For simplicity and without loss of generalicity, we assume that the mixture is composed by two normals (if one has other mixture distribution, usually one can cope with it through the same approache summarized here). We write the mixture distribution as follows. Here we use the exponential twist method.

$$f = \beta_1 g_1 + \beta_2 g_2 \tag{4}$$

where $\beta_1 + \beta_2 = 1$, and $g_{1,2}$ are normal distributions. It yields,

$$M(s) = \beta_1 e^{\sigma_1^2 s^2/2+\mu_1 s} + \beta_2 e^{\sigma_2^2 s^2/2+\mu_2 s} \tag{5}$$

Since $\mu(s) = log(M(t))$, and let $\mu'(s_t) = t$, it yields,

$$t = \frac{\beta_1 e^{\sigma_1^2 s^2/2+\mu_1 s}(\mu_1 + \sigma_1^2 s) + \beta_2 e^{\sigma_2^2 s^2/2+\mu_2 s}(\mu_2 + \sigma_2^2 s)}{\beta_1 e^{\sigma_1^2 s^2/2+\mu_1 s} + \beta_2 e^{\sigma_2^2 s^2/2+\mu_2 s}} \tag{6}$$

Then

$$\beta_1 e^{\sigma_1^2 s^2/2+\mu_1 s}(t - \mu_1 - \sigma_1^2 s) + \beta_2 e^{\sigma_2^2 s^2/2+\mu_2 s}(t - \mu_2 - \sigma_2^2 s) = 0 \tag{7}$$

For simplicity, we just call the left side of equation (7) $F(s_t)$. Namely,

$$F(s_t) = \beta_1 e^{\sigma_1^2 s^2/2+\mu_1 s}(t - \mu_1 - \sigma_1^2 s) + \beta_2 e^{\sigma_2^2 s^2/2+\mu_2 s}(t - \mu_2 - \sigma_2^2 s) \tag{8}$$

Thus, the problem is how to find a special s_t which statisfies $F(s_t) = 0$. That s_t is the optimal parameter in the IS simulation when $\{\mu_1, \sigma_1, \mu_2, \sigma_2, t\}$ are given. There are several ways to realize it. Here we explain how to solve it using the bisection method.
1) firstly, one needs to find two points, say, s_0 and s_1, let $F(s_0) > 0$, and $F(s_1) < 0$.
2) secondly, one needs to compute the midpoint between s_0 and s_1, say, s_{mid}, namely, $s_{mid} = \frac{s_0+s_1}{2}$.
3) thirdly, let $s_0 = s_{mid}$ when $F(s_{mid}) > 0$ holds, or let $s_1 = s_{mid}$ when $F(s_{mid}) < 0$ holds.
4) repeat step 2) and 3) till the predeterminded accuracy of the solution is reached.

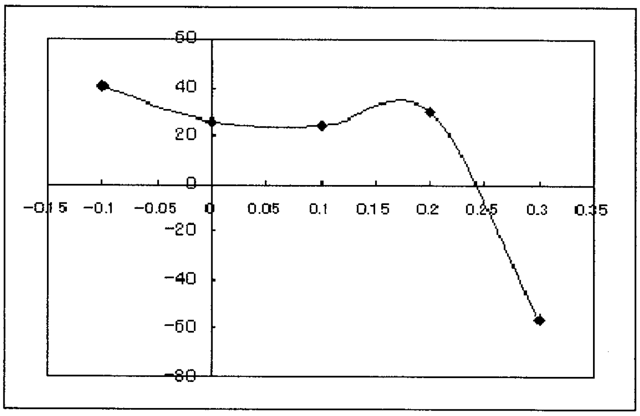


Figure 6: Solution between -0.1 and 0.3

On the other hand, the optimal density function $f^*(x)$ turns out to be

$$\begin{aligned} f^*(x) &= e^{s_t x-\mu(s_t)} f \\ &= e^{s_t x-\mu(s_t)}\left(\beta_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \beta_2 \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) \\ &= \beta_1 e^{-\mu(s_t)+\frac{(\mu_1+\sigma_1^2 s_t)^2-\mu_1^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-(\mu_1+\sigma_1^2 s_t))^2}{2\sigma_1^2}} \\ &+ \beta_2 e^{-\mu(s_t)+\frac{(\mu_1+\sigma_2^2 s_t)^2-\mu_2^2}{2\sigma_2^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-(\mu_2+\sigma_2^2 s_t))^2}{2\sigma_2^2}} \end{aligned}$$

It means that optimal density function of the mixture distribution is still a mixture distribution of normals. The generated random numbers are distributed by f^* in the IS simulations.

Here we give an example, where $\beta_1 = 0.2, \beta_2 = 0.8, \mu_1 = 0, \mu_2 = -1, \sigma_1 = 10, \sigma_2 = 5$, and $t = 25$, corresponding $\hat{p}_t = 0.0012420158$.

Seen from Fig.6, let $s_0 = -0.01$, and $s_1 = 0.30$, then we have $F(-0.01) = 26.46364901 > 0$, and $F(0.3) = -56.24529597 < 0$, respectively. We use the above-explained bisection method to search the optimal s_t. As shown in Fig. 6, the solution is between -0.1 and 0.3. And finally we get $s_t = 0.280214385$ while $F(s_t) = -1.10134E^{-13}$.

Actually it can be easily extended to a mixture distribution which is constructed by n normals. Since $F(-\infty) \to +\infty > 0$, and $F(+\infty) \to -\infty < 0$, and $F(s_t)$ is a continuous function, then the optimal s_t that satisfies $F(s_t) \approx 0$ can be obtained after bisection algorithm has been repeated enough times. Meanwhile, the IS density function f^* still remains as a mixture distribution of n normals, even though the mixture distribution has n normal components.

Table 3. Estimates of p_t and τ

	p_t	τ
1	0.001247185	234.0255
2	0.001244696	234.0904
3	0.001258913	228.5909
4	0.001233208	238.0388
5	0.001255329	232.2330
	...	...
E(·)	0.001243067	233.2091

The estimated tail probability $\hat{p}_t$ and efficiency τ are shown in Table 3. Seen from the Table 3, the tail probability has been precisely estimated by the IS method. Besides, it shows that the IS method is much more efficient than the standard Monte Carlo simulation.

Another example we give here is the case of Data A. We have got its mixture distribution by the GA already. It is a mixture distribution with three nomal components. We apply the IS method to it to estimate the tail probability as we explain above. The results are summarized in Table 4 and Table 5.

Table 4. Estimates of $\hat{p}_t$ and τ

t	2
s_t	0.702334
$F(s_t)$	-1.11022E^{-15}
$E(\hat{p}_t)$	0.046062984
$E(\tau)$	4.5043000

Table 5. Estimates of $\hat{p}_t$ and τ

t	3
s_t	0.80927251
$F(s_t)$	2.22045E^{-16}
$E(\hat{p}_t)$	0.012730215
$E(\tau)$	13.81202124

Seen from Table 5 and 6, the optimal parameter s_ts in the IS simulation are successfully found by the bisection method, and the tail probabilities (odds of rare events) have been accurately estimated by the IS method. Besides, the advantage of the IS simulation to the standard Monte Carlo simulation is confirmed here as well.

5 Conclusion

In this paper we have shown how to use a mixture distribution to approximate a probability distribution function based upon the observed data sets. As to optimize the parameters (combinational weights and the distributional parameters) in the mixture distribution, we have proposed to apply Genetic Algorithm to it. Furthermore, as to do the tail distribution analysis, we have explained how to find the optimal parameter in the Importance Sampling simulation using bisection algorithm, and have shown that the optimal solution is guaranteed if the mixture distribution is constructed by several normals. Through the numerical experiements on the real stock market data sets, it has been confirmed that Genetic Algorithm is powerful tool in serch of the optimal parameters in the mixture dsitributions, and the efficiency and accuracy have been remarkably improved by using the Importance Sampling simulation in the tail distribution analysis.

For future works, it is necessary to extend the method to other fields and to include more complicated radical distribution functions. Further researches will be done by the authors.

Acknowledgments

This research was partly supported by the Japan Society for the Promption of Science under the grant number (B) 15310120, (C) 19510164, and the authors would like to thank the organization.

References

[1] V. Akgiray, G. G. Booth, "The stable-law model of stock returns", Journal of Business Economy Statistics, 6, pp.51-57,1998.

[2] R. K. Bhattacharjya, "Optimal design of unit hydrographs using probability distribution and genetic algorithm", Sadhana, vol.29, Part 5, pp.499-508, 2004.

[3] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu and X. Yang, "Long-time dependency and heavy-tail modeling for traffic data", IEEE Signal Process.Mag.,vol.19, no.3, pp.14-27, 2002.

[4] P.Dutre, F.Suykens and Y.Wilems, "Optimizaed Monte Carlo path generation using Genetic Algorithm", Report CW267, Department of computer science, Katholieke Universiteit Leuven, 1998.

[5] M.Fernandes, "Financial crashes as endogenous jumps:estimation, testing and forecasting", Journal of Economic Dynamics & Control, vol.30, pp.111-141, 2006.

[6] P.Glasserman,P.Heiderberger and P.Shahabuddin, "Asymptotically optimal importance sampling and stratification for path-dependent options", Journalof Mathematical Finance, vol.9, pp.117-152, 1999.

[7] P.Glasserman,P.Heiderberger and P.Shahabuddin, "Variance reduction techniques for estimating Value-at-Risk", Management Science, vol.46, no.10, pp.1349-1364,2000.

[8] D. Goldberg, Genetic Algorithm: in Search, Optimization, and Machine Learning, Addison-Wesley Press, 1989.

[9] P. Heiderberger, "Fast estimation of rare events in queueing and reliability models", ACM Transaction on Modeling and Comput. Simulation, vol.5, pp.43-85, 1995.

[10] P. Jorion, Value at Risk: The New Benchmark for Controlling Market Risk, McGraw-Hill, 1997.

[11] S. M. Kay and S. Saha, "Mean likelihood frequency estimation", IEEE Trans., Signal Processing, vol.48, no.7, pp.1937-1946, 2000.

[12] J. Lyon, B. Barber and C. Tsai, "Improved methods for tests of long-run abnormal stock returns", Journal of Finance, vol.54, pp.165-201, 1998.

[13] M. Song and R. M. Haralick, "Optimal multidimensional quantization for pattern recognition", Proc. SPIE vol.4875, pp.15-30, Second International Conference on Image and Graphics, 2004.

[14] K. Tan and S. Tokinaga, "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.

[15] K. Tan and S. Tokinaga, "The design of multistage fuzzy inference system with small number of rules based upon the optimization of rules by using GA". IEICE Trans.Fundamentals, vol.E82-A, no.9, pp.1865-1873, 1999.

[16] K. Tan, F. Chu and S.Tokinaga, "Estimation of tail distribution of performance evaluation functions based on the Importance Sampling in stochastic models described by variables including jump diffusion processes and its applications (in Japanese) ", IEICE Trans.Fundamentals,to appear, 2006.

[17] K. Tan and S. Tokinaga, "Identifying returns distribution by using mixture distribution optimized by Genetic Alogrithm", Proceeding of NOLTA2006, pp.119-122, 2006.

[18] S. Tokinaga and N. Takagi, "Decomposition of surface data into fractal signals based on mean likelihood and importance sampling and its applications to feature extraction", IEICE Trans. Fundamentals, vol.E88-A, no.7, pp.1946-1956, 2005.

[19] A. Wolfgang, N. N. Ahmad, S. Chen and L. Hanzo,"Genetic alogorithm assisted error probability optimisation for beamfoming", Electronics letters, vol. 40, no. 5,2004.

Kangrong Tan (Professor, Faculty of Economics, Kurume University)
Shozo Tokinaga (Professor, Graduate School of Economics, Kyushu University)