Multiple Suggestions for Interactive SDE Estimation

Ryoji Fukuda
rfukuda@oita-u.ac.jp
Faculty of Engineering,
Oita University,
Japan

Hiromi Yokoyama
v12e6019@oita-u.ac.jp
Faculty of Engineering,
Oita University,
Japan

Abstract: *We are developing educational software to estimate the parameters of stochastic differential equations (SDE) using a single set of time series data. Our target equation is a linear SDE with constant coefficients, which is determined by four real parameters. In our previous version, for a single set of data, we obtained a single set of estimated parameters and suggestions to change them. In this study, we propose a method to give several sets of estimated parameters, assuming several situations for the first estimation. Then, we will be able to find a closer set of parameters, and the interactive estimation will be improved.*

1. Introduction

A stochastic differential equation (SDE) is a differential equation in which one or more terms represent stochastic processes. SDEs are used in various fields; their application to mathematical finance is well known. Our target equation is a linear SDE with constant coefficients. This equation is determined by using two linear functions; in other words, four real constants control this equation. This equation is simple and many experts believe that it is easily understandable. However, it may be difficult for them to explain the role of each parameter to laymen.

One of the main uses of our educational software is to understand the roles of parameters by using simulated modifications. First, our system estimates the values of SDE parameters with respect to a given set of time series data [1]. Then, a user changes the values of the parameters, which leads to the corresponding sample paths being created. If these values differ significantly from the correct values, the user will be able to perceive the difference by comparing the sample paths several times. First, a user may realize that the given parameters are not sufficiently accurate. After extensive trial and error, we expect the user to learn the roles of the parameters and to improve them accordingly. In our previous version of the software, the estimated parameters were highly inaccurate and the user was unable to improve them. The objective of this study is to facilitate the improvement of the parameters.

We use an evaluation function for the correctness of the parameters in order to estimate them. The evaluation value is obtained from a generalized Choquet integral with respect to a two-additive measure. We define several types of feature values that can be used for verifying the correctness of the parameters. The two- additive measure is defined on a set of feature value types, and the integrand is the corresponding feature value for each type. First, we try to improve these feature values and the generalized Choquet integral. For this integral we use Dombi's t-norm, which is a binary operator defined on [0,1]. This operator is a generalization of the minimum operator. In our new estimation, we also use Dombi's t-conorm, which is a generalization of the maximum operator. Assume that these binary operators are defined on a set of membership functions with respect to fuzzy sets. Then, a t-norm corresponds to the intersection and a t-conorm corresponds to the union.

Two-additive measures are obtained from linear regressions. All feature values are obtained by a modification of the process and four parameters. If the corresponding stochastic process is a solution of some SDE, the evaluation values should be small when the given parameters and correct SDE parameters are similar. For obtaining the training data, the SDE parameters are selected randomly, and the corresponding sample path is obtained using random numbers; subsequently, we obtain random parameters and the distance from the correct SDE parameters. Moreover, we control the distance so that it does not exceed D. In the previous version, we consider a single value of D, whereas in our new method, we consider several values of D; accordingly, we obtain several evaluation functions. Thus, we obtain several candidates for estimating the parameters.

2. Stochastic Differential Equation and the System

Our target SDE is a linear SDE with constant coefficients, and the objective of developing our system is to use it for estimating the four parameters of the SDE. A set of time series data is used for estimating the parameters. In this section, we describe our system in detail.

2.1 Stochastic Differential Equation

In this study, we consider an SDE of the form

$$X_t - X_s = \int_s^t (aX_r + b)dr + \int_s^t (c X_r + d) dB_r , \quad (1)$$

where a, b, c and d are real constants, dr denotes the Lebesgue integral, and dB_r denotes the Ito stochastic integral with respect to the Brownian motion process $B(t)$. Equation (1) can be expressed in a condensed form as

$$dX_t = (aX_t + b)dt + (c X_t + d) dB_t .$$

This equation roughly indicates that the mean value of the increment in X_t is proportional to with $(aX_t + b)dt$, and the corresponding variance is $(c X_t + d)^2 dt$. It is well known that equation (1) has a unique strong solution [2] given by

$$X(t) = U(t)\left(X(0) + \int_0^t \frac{a - c*d}{U(s)} ds + \int_0^t \frac{c}{U(s)} dB(s)\right),$$

$$U(t) = \exp\left(\int_0^t \left(b - \frac{1}{2} d^2\right) ds + \int_0^t d\, dB(s)\right) . \quad (2)$$

A simulated modification is given by the above solution. The distribution of the solution is determined by the parameters a, b, c and d; however, explicit values contain random elements. Then we need to create the modifications several times for comparison with the given data. Assume that the given data set is a sample path of the solution (2) of the SDE in (1), and that the corresponding parameters are a, b, c and d. The graphs for the given data and simulated modification may have some common features, and the modification may be very similar to the given data after several iterations of creation.

2.2 Outline of the System

By clicking the “Draw Original” button, four parameters are randomly defined and a sample path is obtained as a simulated modification (Figure 2.1). We assume that this is a given set of time series. At the same time, the system estimates the four parameters, and these values are listed in the lower-left text boxes. For this estimation we have nine estimation methods. Here, we use a generic algorithm for this estimation. The estimation essentially depends on the evaluation function. Our evaluation functions are expressed as generalized Choquet integrals with respect to a two-additive measure. The integrands are feature values obtained from a set of time series data and the four parameters. In the previous version, we fixed a single two-additive measure. In this version, we prepared 9 two-additive measures according to the area of the distance between a target parameters and the correct parameters. In section 4, we will provide the relevant details. Thus, we obtain nine sets of estimated parameters. After considering several graphs of their modifications, we obtain good parameters and bad parameters. These may be valid indicators of the fine estimation of the parameters.

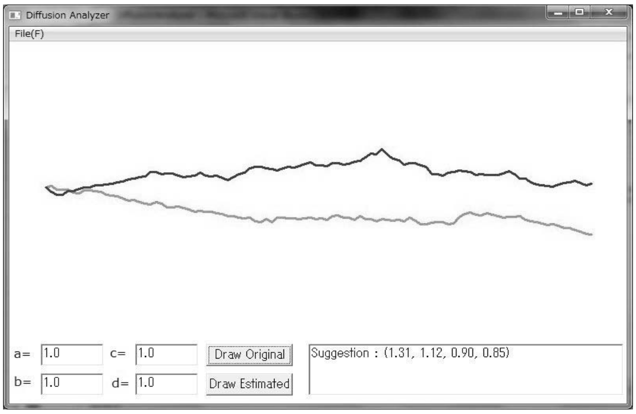

Figure 2.1 Screenshot of the system

3. Restudy of Features and Choquet Integral

In our previous system [1], the evaluation function for the estimation of the parameters is expressed as a generalized Choquet integral. The estimated parameters are generally not adequate enough, and this section discusses restudies of several parts of them.

3.1 Feature values.

Let $\{X_t\}$ be a modification of the solution of the SDE in (1) with some unknown parameters. We observe a set of time series data, $\{X_{t_k}\}_{k=1,2,...}$. This SDE is often expressed as

$$dX_t = (aX_t + b)\,dt + (cX_t + d)\,dBt \,. \tag{3}$$

Assume that (3) is strictly true in case dt, dX_t and dB_t is replaced by $t_k - t_{k-1}$, $X(t_k) - X(t_{k-1})$, and $B(t_k) - B(t_{k-1})$, respectively, and define

$$\gamma_k = \frac{B(t_k) - B(t_{k-1})}{\sqrt{t_k - t_{k-1}}} = \frac{X(t_k) - X(t_{k-1}) - (aX(t_{k-1}) + b)(t_k - t_{k-1})}{(cX(t_{k-1}) + d)\sqrt{t_k - t_{k-1}}}$$

Then, the sequence $\{\gamma_k\}$ is a independent and identically distributed and distributions is $N(0,1)$ (iid $N(0,1)$). Using Ito's formula, the process X_t^2 is also expressed by the same type integrals, and we obtain another approximating sequence $\{\gamma'_k\}$ of iid $N(0,1)$. Using four properties of iid $N(0,1)$, we create eight feature values in the previous version of the system.

In the new version, we no not use $\{\gamma'_k\}$ since there are too many neglected terms. In the estimation of the evaluation functions, we use differentials of feature values with respect to SDE parameters. In the previous version, two feature values take integer values and we cannot define their differentials. We do not use these feature values. We use the following seven feature values in the new system.

1. $y_1 = \left| \frac{1}{n} \sum_{k=1}^{n} \gamma_k \right|^2$
2. $y_2 = \left| \frac{1}{n/2} \sum_{k=1}^{n/2} \gamma_k \right|^2 + \left| \frac{1}{n/2} \sum_{k=n/2+1}^{n} \gamma_k \right|^2$

3. $y_3=\left|\frac{1}{n/2}\sum_{k=1}^{n/2}(\gamma_k-\bar{\gamma}_F)^2+-\frac{1}{n/2}\sum_{k=n/2+1}^{n}(\gamma_k-\bar{\gamma}_L)^2\right|$

4. $y_4=\left|\frac{1}{n/2}\left(\sum_{k=1}^{n/4}(\gamma_k-\bar{\gamma}_B)^2+\sum_{k=3n/4+1}^{n}(\gamma_k-\bar{\gamma}_B)^2\right)-\frac{1}{n/2}\sum_{k=n/4+1}^{3n/4}(\gamma_k-\bar{\gamma}_C)^2\right|$

5. $y_5=\left|\frac{1}{n/2}\sum_{k=1}^{n/2}(\gamma_{2\mathrm{k}}-\bar{\gamma}_E)(\gamma_{2\mathrm{k}-1}-\bar{\gamma}_O)\right|$

6. $y_6=\left|\frac{1}{n/2}\sum_{k=1}^{n/2}(\gamma_k-\bar{\gamma}_F)(\gamma_{k+n/2}-\bar{\gamma}_L)\right|$

7. $y_7=\left|\left(\frac{1}{n/2}\sum_{k=1}^{n/2}(\gamma_k-\bar{\gamma})^2\right)-1\right|$

Here, the following definitions hold:

$$\bar{\gamma}=\frac{1}{n}\sum_{k=1}^{n}\gamma_k,\quad \bar{\gamma}_F=\frac{1}{n/2}\sum_{k=1}^{n/2}\gamma_k,\quad \bar{\gamma}_L=\frac{1}{n/2}\sum_{k=n/2+1}^{n}\gamma_k,$$

$$\bar{\gamma}_B=\frac{1}{n/2}\left(\sum_{k=1}^{n/4}\gamma_k+\sum_{k=3n/4+1}^{n}\gamma_k\right),\quad \bar{\gamma}_C=\frac{1}{n/2}\sum_{k=n/4+1}^{3n/4}\gamma_k,$$

$$\bar{\gamma}_E=\frac{1}{n/2}\sum_{k=1}^{n/2}\gamma_{2k},\quad \bar{\gamma}_O=\frac{1}{n/2}\sum_{k=1}^{n/2}\gamma_{2k-1}.$$

y_1, y_3 and y_5 are also used in our previous version. All values converge to 0 as n tends to infinity if "iid and $N(0,1)$" are true (using the property "the mean values are zero" for y_1 and $y_{2,}$, "identically distributed" for y_3 and y_4, "independence" for y_5 and y_6, and "the variances are 1" for y_7).

3.2 Generalization of Choquet Integral

Let μ be a two-additive measure is given by

$$\mu(A)=\sum_{x\in A}\mu_x+\sum_{\{x,y\}\subset A}\nu_{x,y},$$

where μ_x and $\nu_{x,y}$ ($x,y\in A$) are real constants. Thus, two-additive measure on an n-point set is defined by $n+n(n-1)/2$ values. The generalized Choquet integral of a function f ($A\rightarrow[0,1]$) is given by

$$\int_A f\,d\mu=\sum_{x\in A}f(x)\mu_x+\sum_{\{x,y\}\subset A}f(x)\otimes f(y)\nu_{x,y},$$

where $\otimes$ is Dombi's t-norm [3] defined by

$$x\otimes y=\frac{1}{((1/x-1)^{\lambda}+(1/y-1)^{\lambda})^{1/\lambda}}$$

(in this case $\lambda=2.5$; see [1]).

Let A be a finite set and a,b be elements of A. Let m be a set function defined by

$$m(B)=\begin{cases}\alpha & \text{if } (a\in B \text{ and } b\in B^c) \text{ or } (a\in B^c \text{ and } b\in B)\\ 0 & \text{otherwise}\end{cases}.$$

While this is a simple and basic set function, it is not a two-additive measure. Generally, the Choquet integral is defined for a set function with the value correspond to the empty set is zero, and m clearly satisfies this property. Therefore, we can consider the Choquet integral of a function f defined on A, which is given by

$$\begin{aligned}\int_A f\, dm &= \int_0^\infty m(\{x : f(x) > r\})dr \\ &= (f(a) \vee f(b) - f(a) \wedge f(b))\alpha\end{aligned}.$$

Our Choquet integral is generalized by replacing $\wedge$ by a t-norm. Thus, we consider that the use of the t-conorm term is useful. We generalize the Choquet integral as follows:

$$\int_A f\, d\mu = \sum_{x\in A} f(x)\mu_x + \sum_{\{x,y\}\subset A} f(x)\otimes f(y)\nu^{(t)}_{x,y} + \sum_{\{x,y\}\subset A} f(x)\oplus f(y)\nu^{(c)}_{x,y}, \qquad (4)$$

where, $\oplus$ is the t-conorm defined by Dombi's t-norm, that is, the binary operator $\oplus$ on [0,1] is defined by

$$x\oplus y = 1-(1-x)\otimes(1-y).$$

The integral defined by (4) is a generalized Choquet integral with respect to the two-additive measure, and the family of real numbers $\{\mu_x, \nu^{(t)}_{x,y}, \nu^{(c)}_{x,y} : x, y\in A\}$ determines the integral.

3.3 Evaluation of Distance of Parameters

Let A be the set of all feature value types. For the previous version, A consists of eight elements, while for the new version there are seven elements in it. Using a set of time series data, we obtain the corresponding feature values. That is, a set of time series data defines a function on A.

For a function f, the generalized Choquet integral (4) is a linear combination of

$$f(x), f(x)\otimes f(y), f(x)\oplus f(y), x, y\in A.$$

We approximate the parameters $\{\mu_x, \nu^{(t)}_{x,y}, \nu^{(c)}_{x,y} : x, y\in A\}$ using linear regressions. Training data are obtained using random numbers as follows.

- T-1. Parameters a, b, c and d are given using $N(0,1)$ random numbers.
- T-2. Create a simulated modification with respect to these parameters.
- T-3. The random distance value r is given: r is uniformly distributed on $[0, D]$.
- T-4. Another set of random parameters a', b', c', d' are given, and these satisfy the relation $\sqrt{(a-a')^2+(b-b')^2+(c-c')^2+(d-d')^2} = r$.
- T-5. Calculate the feature values by using the parameters a', b', c' and d', and extend them using t-norm and t-conorm.
- T-6. Iterate T-3 to T-5, 20 times (base iteration).
- T-7. Iterate T-1 to T-6, 20 times (local iteration).

Using this procedure, we obtain 400 sets of training data. Each set consists of a 36- or 49-dimensional vector $\vec{y}^{(p)} = \{(y^{(p)}_j, y^{(p)}_j \otimes y^{(p)}_k) : j, k \le 8\}$, $\vec{y} = \{(y_j, y_j\otimes y_k, y_j \oplus y_k) : j, k \le 7\}$ (these were defined in subsection 3.1), and r (distance of two parameter sets), where $y^{(p)}_j$ ($j \le 8$) are feature values used in the previous version. Then, we obtain the coefficient vectors $\vec{a}'$ and $\vec{a}$ and the constants b' and b that minimize

$$\sum_i |r_i - (\vec{a}'\cdot\vec{y}^{(p)} + b')|^2, \quad \sum_i |r_i - (\vec{a}\cdot\vec{y} + b)|^2.$$

The coefficient vectors determine the integral, and the estimation of the parameters involves the use of these integrals. For the linear regressions, we use pseudo-inverse in the case where some eigenvalues are almost equal to zero.

The following steps are performed for evaluating evaluating coefficients.

C-1. Using the steps T-1 to T-7, we obtain a set of training data. to create coefficient

C-2 Using the data set, we obtain the coefficient vectors and constant values.

C-3. Using the steps T-1 to T-7 again, we obtain another data set.

C-4. Using the coefficient vectors and constant values obtained in C-2 and the set of data obtained in C-3, we calculate the square errors.

	Old	New
Feature dim.	8	7
Coeff. dim.	36	49
Base iteration	20	
Local iteration	20	
D (max dist)	1, 2, 3, 4, 5, 7, 10, 15, 20	

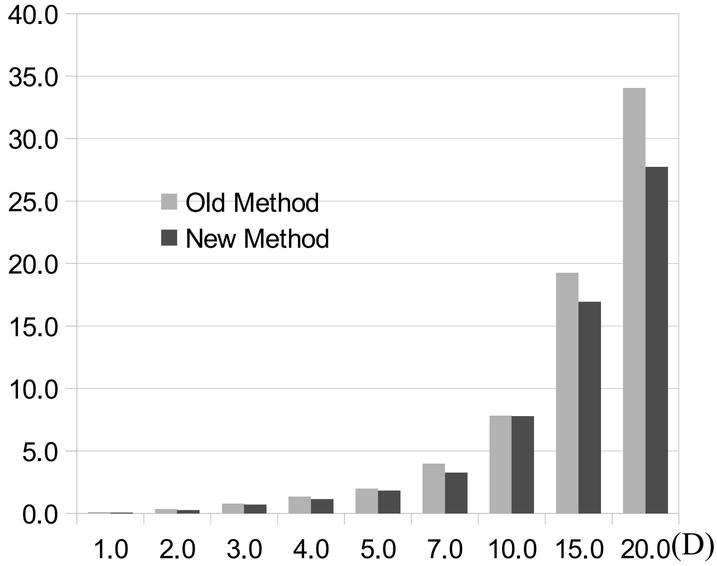

Figure 3.1 Square errors for distance approximations

4. Multiple Suggestions

Using the generalized Choquet integral given in the previous section, we estimate the SDE parameters by using the following generic algorithm [1].

1. The initial values of the SDE parameters are obtained by using the methods explained in Section 5. This vector is copied to all 100 gene vectors.
2. N(0,0.1) (independent) random numbers are added to every coefficient of all gene vectors.
3. The gene vectors are sorted in an ascending order of the evaluation functions discussed in the previous subsection.
4. Recombination (we will explain this step below) is performed according to the sorting result.
5. Parameters obtained after 50 iterations of 3 and 4 are considered to be the final estimated parameters.

The First three gene vectors are left intact. The remaining gene vectors are divided into three groups of the same size, and replaced by the following vectors:

0. Let $\{\vec{v}_i\}_{i=1}^{32}$ be the first 32 gene vectors, including the first 3 vectors.
1. Genes in the first group are replaced by improved vectors using the Newton method.
2. Genes in the second group are replaced by improved vectors using the gradient method. We use a random variable as the coefficient of the gradient.
3. Genes in the third group are replaced by $\{\vec{v}\}_{i=1}^{32}$, and then these values are modified using N(0,0.1) random numbers.

In the previous version, two feature values are given by counting. Therefore, these are not differentiable (with respect to the parameters) and are ignored in the steps involving the Newton method and gradient method. The new feature values are differentiable and there are no ignored components in this analysis.

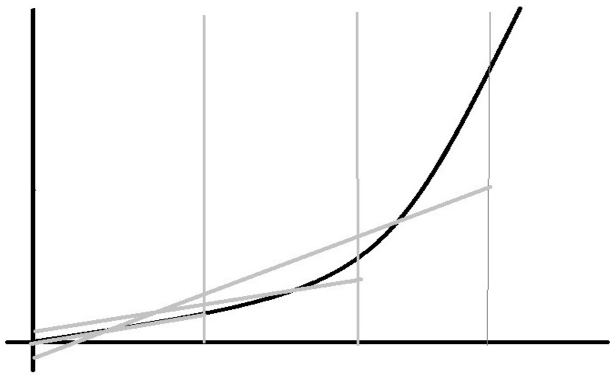

Figure 4.1 Linear Regressions

Generally, linear regression approximates a given function by a linear function. The function and the corresponding performance will depend on the data distribution. The example in Figure 4.1 shows that a regression line with respect to a small area does not generally approximate well in a wider area. Our approximation of the parameters by a generalized Choquet integral involves the use of linear regression, and this ability may depend on the distance of the parameters from the correct ones. We have to prepare several approximated functions for obtaining the distance of the parameters.

4.1 Approximations According to Error Sizes

We should change the regression functions according to the neighborhood sizes . We prepare nine areas for error size. By using the following procedure, we obtain nine regression functions to approximate the distance from the correct parameters.

1. We use 9 D-values:(1.0, 2.0, 3.0, 4.0, 5.0, 7.0, 10.0, 15.0, 20.0). ([0,D] is the distance interval between two sets of SDE parameters. This was explained in T-3 in subsection 3.3.)
2. For each D-value, using the steps T-1 to T-7 we obtain 400 teacher data.
3. Linear regression is used to calculate the coefficients.

For the evaluation, we obtain square errors by the following steps.

4. For the fixed D, we select a comparative parameter D' among nine D-values not less than D.
5. Using the steps T-1 to T-7 with D', we obtain the feature vectors and parameters, for which the distance from the correct one is not more than D'.

The graphs in Figure 4.2 show errors for fixed D'(= 5.0, 20.0). The errors takes minimal values when $D=D'$

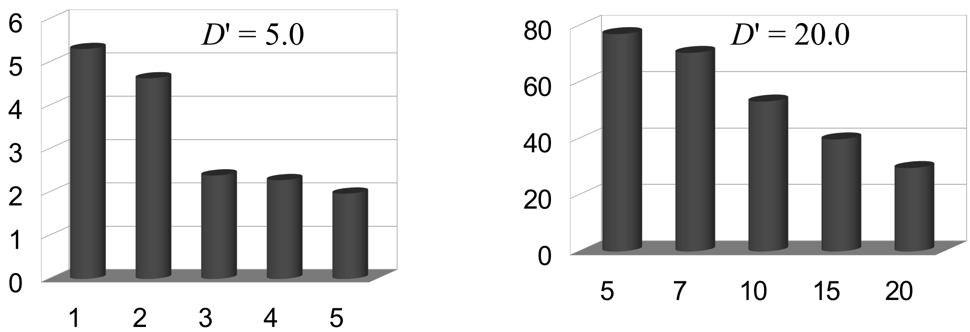

Figure 4.2 Squire Errors for fixed Error Areas

4.2 Estimation of Parameters

In the system, for a given set of time series data, we obtain the estimated parameters. In this study, we consider several types of estimation methods. In the previous subsection, we discussed the creation of nine approximation functions for distances of parameters. At first, a rough estimation of the parameters is made, and a genetic algorithm is used to improve the parameters. In this estimation, the system uses one approximated function from among the nine options. This estimation depends on the selected function. We can select one of the function after trying all of them several times. In other words, assume that we can select the optimal parameters from the graph of simulated sample paths. We can improve the estimation, unlike the case where a single approximated function is used. Figure 4.3 shows the estimation result for 10 trials. The differences between distances are not very large for one set of original data. However similarities of graph shapes for simulated modifications of estimated parameters are quite different from each other (see Figure 4.4 for an example), and the extent of similarity does not change for one set of estimated parameters.

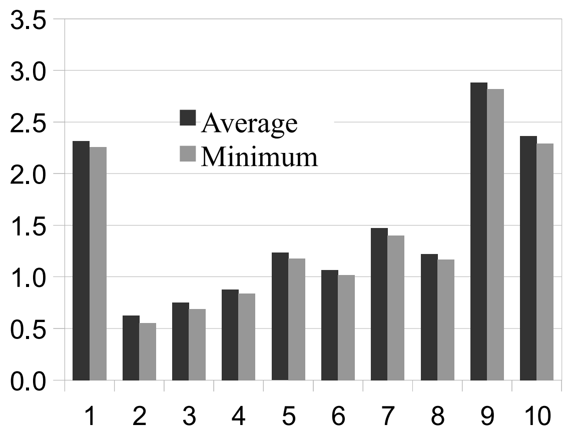

Figure 4.3 Average and minimum values of distances

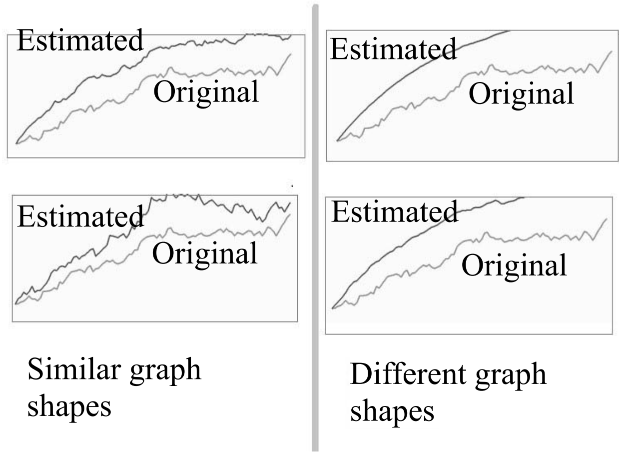

Figure 4.4 Graph shapes

5. Conclusion

In this study, we analyzed our estimation methods. It is important is to approximate distance of parameters. Our first estimation of the parameters is not always good, and when a wrong approximation area is used, the corresponding estimations are incorrect. We used several error distance areas and increase the choice of estimated parameters for the stochastic differential equation. Our next problem is the automatic selection of the approximation function.

References

[1] R. Fukuda, T. Abe, Interactive Estimator for Stochastic Differential Equation, Proceedings of the 16th Asian Technology Conference in Mathematics, 2011

[2] B. Oksendal, Stochastic Differential Equations: An Introduction with Applications, Springer, 1985

[3] J. Dombi, A General Class of Fuzzy Operators, the DeMorgan Class of Fuzzy Operators and Fuzziness Measures Induced by Fuzzy Operators, Fuzzy Sets and Systems, 8, 149-163, 1982