Levenberg-Marquardt Parameter Estimation of Stochastic Logistic Model

Haliza Abd.Rahman[1*],Arifah Bahar[1],Mohd.Khairul Bazli[1],Norhayati Rosli[1],Madihah Salleh[2] and Gerhard-Wilhelm Weber[3].
[1]Department of Mathematics, Faculty of Science, UTM,81310, Skudai, Johor, Malaysia.
[2] Department of Biotechnology Industry, Faculty of Biosciences and Bioengineering, UTM,81310, Skudai, Johor,Malaysia.
[3]Middle East Technical University, Institute of Applied Mathematics, 06531 Ankara, Turkey.
Email:haliza.ar@fs.utm.my

Abstract

In this paper, we estimated the drift and diffusion parameters of the stochastic logistic models for the growth of Clostridium Acetobutylicum P262 using Levenberg Marquardt optimization method. The parameters had been estimated for three different substrates. Milstein scheme had been used to approximate the solution of stochastic differential equation. The values of Mean Square Errors (M.S.E) of the stochastic logistic models indicated a good fit.

Keywords

Ito Differential Equation, Stochastic Logistic Model, Levenberg Marquardt, Milstein Scheme, Fermentation.

INTRODUCTION

Many physical phenomena can be better presented and understood via mathematical modeling. Wide literatures on mathematical modeling of physical systems may be found with deterministic modeling particularly by deterministic differential equation whereby the element of noise is not considered. It may present optimum situations and can be improved by introducing stochastic element since in reality many phenomena in nature are affected by stochastic noise therefore *stochastic differential equations* (SDE) may be required. Some of the fields which apply SDE are finance, population dynamics, hydrology, environmetric and biometry [1].

The stochastic modelling can be considered from the corresponding Ito or Stratonovich differential equations (SDEs) or from the associated Kolmogorov (Fokker-Planck and backward) differential equations [2]. Both represent extrinsic stochasticity whereby the stochasticity is introduced by incorporating multiplicative or additive stochastic terms into the differential equation. Extrinsic stochasticity is due to random variation of one or more environmental or external factors such as temperature or concentration of reactant species whereas intrinsic stochasticity is inherent to the system, arising due to the relatively small number of reactant molecules. Intrinsic stochasticity can be described by a chemical master equation developed by Gillespie in 1977 [3]. The general form of the Ito SDE is as the following,

$$dx(t) = f(x,t) + g(x,t)dW(t) \qquad (1)$$

$x(t)$ is the state of the physical system at time t

$f(x,t)$ is the deterministic or average drift term

$g(x,t)$ is the diffusive term

$dW(t)$ is the Brownian noise

Equation (1) is a one-dimensional stochastic process W perturbing x.

LITERATURE REVIEWS

Most of the research in biotechnological areas employ deterministic modeling of systems. For example, in investigating the feasibility of sago strach as carbon source of Acetone-Buthanol Ethanol (ABE) fermentation, deterministic logistic differential equation was used to model the growth of strain Clostridium Acetobutylicum P262 [4]. This however is inadequate since stochastic model would offer a more realistic representation of systems compared to deterministic since system such as cell growth is subject to random fluctuation whether intrinsic or extrinsic. [5] incorporate stochasticity into power law logistic model and studied its properties. Here, we remodel cell growth of C.Acetobutylicum P262 using stochastic logistic model based on [5] but prior to that, the parameters had to be estimated from data. In SDEs, no method of parameter estimation had been specified, however in modeling stochasticity of biological process parameter estimation is a nontrivial task. In this section, we will discuss some reviews of literature in estimating parameters of stochastic differential equations. Some previous works employed methods such as maximum likelihood method [1],[2], methods of moment [6], filtering (e.g extended Kalman filter) [7] and non-linear least squares[8]. In this paper we estimated the parameters of the logistic equation by applying the optimisation method for non -linear least squares Levenberg-Marquardt method (LM method). The L-Marquardt had been widely used to estimate parameters in deterministic models since this method serves as a fast and convenient method, however few literatures are found in stochastic models. Milstein numerical scheme is used to approximate the solution of SDE since this scheme is considered more efficient than Euler Maruyama [1]. The LM method doesn't require the availability of the transition probability and has been applied to estimate the parameters of stochastic model in polymer rheology [8] and is the modified version of the Gauss-Newton method with the simplified form of the Hessian matrix [9]. Here we employed the method to estimate the parameters of stochastic logistic equation with the non-constant or multiplicative noise.

PARAMETER ESTIMATION OF STOCHASTIC DIFFERENTIAL EQUATION

Deterministic logistic equation of cell growth

Power law logistic differential equation was used to describe population dynamic has the following form [5],

$$\frac{dN}{dt} = aN^{\xi} - bN^{\eta} \tag{2}$$

where $N-$ population density

$a-$ growth coefficient

$b-$ crowding coefficient

$a,b,\xi,\eta-$ are constants

Letting $\xi = 1$, $a = \mu_{\max}$, $b = \frac{\mu_{\max}}{x_{\max}}$ $N = x(t)$, $\frac{dN}{dt} = x(t)$ and $\eta = s+1$, the model represents the rate of cell growth kinetic,

$$\frac{dx}{dt} = \mu_{\max} x(t) - \frac{\mu_{\max}}{x_{\max}} x^{s+1}(t) \tag{3}$$

s is an index of the inhibitory effect accounts for the deviation of growth from the exponential relationship. For case when s = 0, it will be a complete inhibition of cell growth, s = 1 reduces to logistic model. If s ranges from 0 to 1 it describes a higher degree of inhibition compared to logistic growth [10]. $\mu_{\max}$ is a constant represents the maximum specific growth rate (h^{-1}), x is cell concentration (g/L), $x_{\max}$ is a maximum cell concentration (g/L). The logistic equation, case for s = 1 was utilised to describe the kinetics of several fermentation systems.

Stochastic perturbation to logistic cell growth kinetic

In this part, the deterministic logistic cell growth kinetic model, equation (3) wil be perturbed by extrinsic Brownian white noise through its growth coefficient a. For simplicity equation (3) is rewritten as

$$x'(t) = ax(t) + bx^{s+1}(t) \tag{4}$$

Every element of growth coefficient will be perturbed as the following,

$a \to a + \sigma x^{s}(t)\,dw(t)$ since this is a one dimensional problem, equation (4) becomes,

$$dx(t) = x(t)\left[a + bx^{s}(t)\right]dt + \sigma x^{s+1}(t)\,dw(t) \tag{5}$$

For logistic model, replacing $a = \mu_{\max}$ and $b = \frac{-\mu_{\max}}{x_{\max}}$, the equation becomes,

$$dx(t) = x(t)\left[\mu_{\max} - \frac{\mu_{\max}}{x_{\max}} x(t)\right]dt + \sigma x^{2}(t)\,dw(t) \tag{6}$$

which is of the form of Ito stochastic differential equation with,

$f(x,t) = x(t)\left[\mu_{\max} - \frac{\mu_{\max}}{x_{\max}} x(t)\right]$ is the average drift term.

and $g(x,t) = \sigma x^{2}(t)$ is the diffusion term. By using experimental data it is of interest to estimate the parameter $\mu_{\max}$ and σ.

Milstein approximation

Many of SDE systems do not have known analytical solution, thus solving these systems numerically is necessary. The usual approximation is the Euler Maruyama and Milstein scheme. Euler-Maruyama method has strong order of convergence that is ½ and weak order of convergence 1. Here we will only consider Milstein scheme since the scheme has a higher order, converges with strong order 1 and considered to be more precise than Euler-Maruyama [1]. Considering the Ito SDE in equation (1) on $\left[t_0, T\right]$ for a given discretisation

$t_0 < t_1 < \ldots < t_n < \ldots < t_N = T$, a Milstein approximation is a continuous time stochastic process satisfying the iterative scheme given by,

$$x_{i+1} = x_i + h_i f(x_i) + g(x_i)\Delta W_i + \frac{1}{2} g(x_i) g'(x_i)\left((\Delta W_i)^2 - h_i\right)$$

with initial value x_0. Based on [9] in order to estimate the parameters, equation (6) has to be rearranged in a form of difference quotient,

$$\dot{\bar{x}} = f(x_i) + g(x_i)\frac{\Delta W_i}{h_i} + \frac{1}{2} g(x_i) g'(x_i)\left(\frac{(\Delta W_i)^2}{h_i} - 1\right) \tag{7}$$

The value $\dot{x}_n$ represents difference quotient based on the experimental data. Given $W \sim N(0,t)$ and $\Delta W \sim N(0, t_{i+1} - t_i)$, ΔW can be approximated using the standard normal distribution thus equation (7) becomes

$$\dot{x} = f(x_i) + g(x_i)\frac{z}{h_i} + \frac{1}{2} g(x_i) g'(x_i)(z^2 - 1) \quad (8)$$

Substituting the average drift term and the diffusion term in equation (8) we obtain the following,

$$\dot{x}_i = \mu_{\max} x_i - \frac{\mu_{\max} x_i^2}{x_{\max}} + \sigma x_i^2 \frac{z_i}{\sqrt{h_i}} + \sigma^2 x_i^3 (z_i^2 - 1) \quad (9)$$

APPLICATION

Results and discussions

Investigation of the feasibility of using sago starch as carbon source for solvent fermentation by Clostridium acetobutylicum P262 had been carried out by Salleh, M [4]. Batch fermentation was carried out and one of the experiments involved the investigation of the effect of different inorganic nitrogen source to 5g/l yeast extract using 50g sago starch/l as carbon source. Three different inorganic nitrogen such as ammonium sulphate $(NH_4)_2SO_4$, ammonium chloride (NH_4Cl) and ammonium hydrogen phosphate ($NH_4H_2PO_4$) were added to the substrate and the media was adjusted to pH 6 at 36^o Celcius [11]. 22 experimental data of cell concentration from direct fermentation of sago starch using three different mixtures of inorganic nitrogen source were observed at unequal time interval totaling of 72 hours. The values of $\mu_{\max}$ had been estimated for deterministic logistic model using Levenberg-Marquardt where the solutions had been approximated by fourth order Runge Kutta numerical scheme. The estimated $\mu_{\max}$ for each subtrate are 1.4429, 0.9406 and 0.5056. The values of parameter sigma in drift equation of logistic stochastic model (6) were determined by minimising the cost function via the Levenberg-Marquardt method,

minimise

$$\sum_{i=1}^{N} \left(\dot{x}_i - \left(\mu_{\max} x_i - \frac{\mu_{\max} x_i^2}{x_{\max}} + \sigma x_i^2 \frac{z_i}{\sqrt{h_i}} + \sigma^2 x_i^3 (z_i^2 - 1) \right) \right)^2$$

The simulation of the Brownian noise were generated using the ziggurat method [1] at 50, 100, 500 and 1000 times and the average of the parameter values were obtained. In order to reduce the variance of the simulated normal random numbers for calculating the predicted values, antithetic variates are employed when generating the numbers. The estimated mean parameters are listed as in table 1.

Table 1. The values of the averaged sigma for different number of simulations

σ	YE2	YE3	YE4
N=50	0.3360	0.0009	0
N=100	0.2564	0.0144	0.0127
N=500	0.4808	0.0139	0
N=1000	0.5556	0.0038	0.0032

For modeling purposes only a single value of sigma will be chosen for each substrate therefore the MSE of the stochastic models were calculated using the value of sigma obtained at different number of simulations of the Brownian noise. The simulation of Brownian noise at N = 100 was chosen since it produces the least MSE of the stochastic model (6) for most substrates. Thus the values of sigma from 100 simulations of Brownian noise were chosen for the stochastic model. The parameter values obtained from the combination of parameter estimation of deterministic model and stochastic model are shown in table 2.

Table 2. Parameters of stochastic models for N=100

N=100	YE2	YE3	YE4
$\mu_{\max}$	1.4429	0.9406	0.5056
σ	0.2564	0.0144	0.0127

Figure 2-4 depicts the plot of stochastic models of cell concentration of C. Acetobutylicum P262.

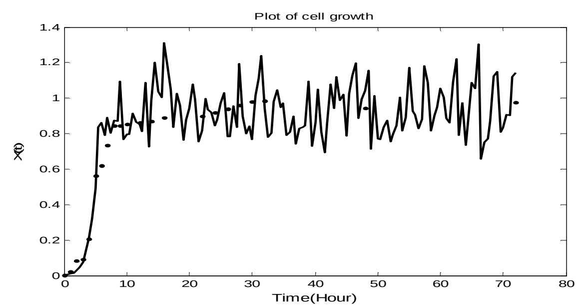

Figure 2. The stochastic logistic model of cell ceoncentration (g/L) versus time (h) of substrates YE2.

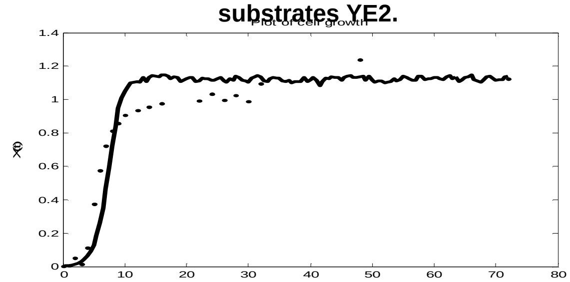

Figure 3. The stochastic logistic model of cell ceoncentration (g/L) versus time (h) of substrates YE3.

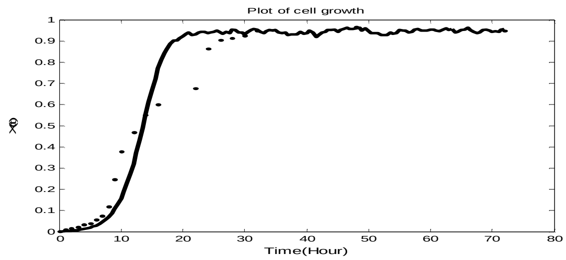

Figure 4. The stochastic logistic model of cell concentration (g/L) versus time (h) of substrates YE4.

PREDICTION QUALITY OF THE MODELS

The performance of the stochastic models for five different substrates are compared with deterministic models using Mean Square (M.S.E). Table 3 below outlined the values of the M.SE for both of the deterministic and stochastic models.

Table 3 Mean Square Error (MSE) of deterministic models and stochastic logistic models used to characterize the prediction quality based on the experimental data.

Substrates	YE2	YE3	YE4
Deterministic	0.0175	0.0124	0.0081
Stochastics	0.0238	0.0194	0.0090

From table 3, it can be seen that the values of M.S.E of both models are approximately the same. The M.S.E are small for both models, thus indicating good fit. However, stochastic logistic models are preferred to model cell growth of Clostridium Acetobutylicum P262 since it gives a more realistic representation of the natural process.

CONCLUSION

In this study, the parameters had not been estimated simultaneously since the values change if different number of data are used to estimate the parameters. We opted to estimate the drift and diffusion parameters separately by estimating the coefficient of the drift term first and the coefficient of the diffusive term next. The L-Marquardt had provided sufficient estimate of the diffusion parameters of the stochastic model by using difference quotient since the solutions obtained are independent of initial values of the parameters. The M.S.E for all five substrates of both deterministic and stochastic models are small and approximately equal and thus describe the adequacy of the stochastic logistic models in modeling of Clostridium Acetobutylicum P262.

ACKNOWLEDGEMENT

This work has been partially supported by Ministry of Higher Eduction (MOHE), Malaysia under Fundamental Research Grant Scheme (FRGS) vot Number 78221 and SLAI scholarship.

REFERENCES

[1] U. Picchini et. al. 2006. Modeling the Euglycemic Hyperinsulinemic Clamp by Stochastic Differential Equations, *Journal of Mathematical Biology*, Vol.53, No.5, pp. 771-796.

[2] R. Gutierrez, et. al. 2008. Trend analysis and computational estimation in a stochastic Rayleigh model : Simulation and application, *Mathematics and Computers in Simulation*, Vol. 77, pp. 209-217.

[3] T.C. Meng, S. Somani and P. Dhar. 2004. Modelling and Simulation of Biological Systems with Stochasticity. *In Silico Biology*. Vol. 4, pp. 293-309.

[4] Salleh, M. 2002. Direct Fermentation of Gelatinised Sago Starch to Solvent (Acetone Butanol-Ethanol) by *Clostridium Acetobutylicum* P262.Phd Thesis. Universiti Putra Malaysia.

[5] Bahar, A. 2005. Applications of Stochastic Differential Equations and Delay Differential Equations in Population Dynamics. PhD Thesis. University of Strathclyde.

[6] Nielsen, J.N, et. al., 2000. Parameters Estimation in Stochastic Differential Equations: An Overview, *Annual Reviews in Control*. Vol. 24, pp. 83-94.

[7] Nielsen, J.N., Madsen, H., 2001, Applying the EKF to stochastic differential equations with level effects. *Automatica*. Vol. 37. No. 1, pp.107-112.

[8] B. P. Lo, et. al. 2008. An algorithm for the estimation of parameters in models with stochastic differential equations, *Chemical Engineering Science*, Vol. 63, pp. 4820-4833.

[9] Taylan, et. al. “Approximation of Stochastic Differential Equations by Additive Models Using Splines and Conic Programming”, *International Journal of Computing Anticipatory Systems* 20-21-22 (2008), ISSN 1373-5411 (at the occasion of CASYS'07, Eighth International Conference on Computing Anticipatory Systems, Liege, Belgium, August 6- 11, 2007) D.M. Dubois (ed.).

[10] Muthuvelayudham,R. and Viruthagiri,T.2007. Optimization and modeling of cellulose protein from Trichoderma reesei Rut C30 using mixed substrate, *African Journal of Biotechnology*, Vol.6, No.1, pp.41-46.

[11] Direct fermentation of gelatinized sago starch to acetone-butanol-ethanol by Clostridium acetobutylicum, *World Journal of Microbiology & Biotechnology* Vol.,17, pp.567-576, 2001