

# Parameter Estimation for the Stochastic SIS Epidemic Model

Jiafeng Pan, Alison Gray, David Greenhalgh, Xuerong Mao  
Department of Mathematics and Statistics,  
University of Strathclyde, Glasgow G1 1XH, U.K.

## Abstract

In this paper we estimate the parameters in the stochastic SIS epidemic model by using pseudo-maximum likelihood estimation (pseudo-MLE) and least squares estimation. We obtain the point estimators and  $100(1 - \alpha)\%$  confidence intervals as well as  $100(1 - \alpha)\%$  joint confidence regions by applying least squares techniques. The pseudo-MLEs have almost the same form as the least squares case. We also obtain the exact as well as the asymptotic  $100(1 - \alpha)\%$  joint confidence regions for the pseudo-MLEs. Computer simulations are performed to illustrate our theory.

**Key words:** Stochastic SIS epidemic model, pseudo-maximum likelihood estimation, least squares estimation, confidence interval, confidence region, asymptotic estimator, logistic equations.

## 1 Introduction

In general parameter estimation in stochastic differential equations (SDEs) is a non-trivial problem [2, 9]. Many SDEs are non-linear, making simpler approaches to estimation impossible to implement. Young [27] reviews parameter estimation methods for continuous time models. Nielsen et al. [21] updates this to include newer methods for discretely observed SDEs. Timmer [26] discusses the relation between Maximum Likelihood Estimators (MLEs) and quasi-MLEs and compares the quasi-MLE approach with the  $\Delta t = \delta t$  approach in simulations. Kristensen et al. [11] considers the stochastic 'grey box' model and presents the approximate MLE approach based on normal approximation and use of the extended Kalman filter and a software package CTSM. Bishwal [2] discusses the asymptotic properties of MLEs and Bayes estimators of real valued drift parameters in SDEs. Parameter estimation for stochastic delay differential equations (SDDEs) in particular has been studied for the last ten years, see e.g. [7, 12, 13, 14, 23, 24, 25].

In this paper we estimate the parameters in the stochastic SIS (susceptible-infected-susceptible) epidemic model as formulated in [6] as an SDE model. This SDE is non-linear, and pseudo-Maximum Likelihood Estimation and least squares parameter estimation will be applied. Our main contribution in this paper is variance estimation. We obtain not only the point estimators but also the interval estimators for parameters and also the joint confidence regions taking the correlation among the parameters and the overall degree of

confidence into account. Also, we investigate the factors which influence the width of the confidence intervals and the area of the confidence regions both analytically and in our simulation examples.

The well-known SIS model is one of the simplest epidemic models and is often used to model diseases for which there is no immunity, including gonorrhoea [8], pneumococcus [15, 17] and tuberculosis.

In our previous paper [6], we derived an Itô SDE SIS model from the deterministic one

$$\begin{cases} dS(t) = [\mu N - \beta S(t)I(t) + \gamma I(t) - \mu S(t)]dt - \sigma S(t)I(t)dB(t), \\ dI(t) = [\beta S(t)I(t) - (\mu + \gamma)I(t)]dt + \sigma S(t)I(t)dB(t) \end{cases} \quad (1.1)$$

with initial values  $S_0 + I_0 = N$ , where  $S_0$  and  $I_0$  are initial numbers of susceptible and infected individuals in a population of size  $N$ .  $S(t)$  and  $I(t)$  are the numbers of susceptible and infected individuals at time  $t$  and  $B(t)$  is a scalar Brownian motion. This model has parameters  $\beta$ ,  $\gamma$ ,  $\mu$  and new parameter  $\sigma$ . Here  $\mu$  is the per capita death rate, and  $\gamma$  is the rate at which infected individuals become cured, so  $1/\gamma$  is the average infectious period. The parameter  $\beta$  is the disease transmission coefficient, so that  $\beta = \lambda/N$  where  $\lambda$  is the per capita disease contact rate. The parameter  $\lambda$  is the average number of adequate contacts of an infective per day. An adequate contact is one which is sufficient for the transmission of an infection if the contact is between a susceptible and an infected individual. Regarding the parameter  $\sigma$ ,  $\sigma^2 dt$  quantifies the size of the variance of the number of potentially infectious contacts that a single infected individual makes with another individual in the small time interval  $[t, t + dt)$ .

As  $S(t) + I(t) = N$ , it is sufficient to study the SDE for  $I(t)$

$$dI(t) = I(t) \left( [\beta N - \mu - \gamma - \beta I(t)]dt + \sigma(N - I(t))dB(t) \right) \quad (1.2)$$

with initial value  $I(0) = I_0 \in (0, N)$ , and in the rest of the paper we will concentrate on this SDE only.

The organization of this paper is as follows: In section 2 we apply the least squares estimation approach to our problem and obtain the point estimators, interval estimators and confidence regions for the model parameters  $\beta$ ,  $\eta = \mu + \gamma$  and  $\sigma^2$ . We consider the cases of parameter estimation for both one sample of data and multiple samples. We also investigate the factors which influence the width of the confidence intervals and the areas of the confidence regions. Simulation examples are given to illustrate our theory. In section 3 we discuss the pseudo-MLE method. We obtain the maximum likelihood estimators and exact and approximate confidence regions, and again consider the case of multiple samples. Also we compare the pseudo-MLEs with the least squares estimators both analytically and in our simulation examples. In section 4 we summarise the findings in the paper and indicate further ongoing work to be reported.

## 2 Least Squares Estimation

We will use the Euler-Maruyama (EM) scheme [18, 19] to approximate the path of the process such that the discretized form of the process can be rearranged as a regression

model. Then the regression theory can immediately be applied to estimate the model parameters. In this section point estimators,  $100(1 - \alpha)\%$  confidence intervals as well as  $100(1 - \alpha)\%$  joint confidence regions will be obtained for our model parameters. Simulation examples will be given to illustrate our theory.

## 2.1 Regression Model

Let  $\{I_k\}_{k=0}^n$  be observations from process (1.2). Given a stepsize  $\Delta t$  and setting  $I_0 = I(0)$ , the EM scheme produces the following discretization over small intervals  $[k\Delta t, (k+1)\Delta t]$

$$I_{k+1} - I_k = I_k(\beta N - \mu - \gamma - \beta I_k)\Delta t + \sigma(N - I_k)I_k\Delta W_k, \quad (2.1)$$

where  $\Delta W_k = W_{k+1} - W_k$ .

Equation (2.1) can be rewritten as

$$y_{k+1} = \eta u_{k+1} + c + \sigma Z_{k+1}, \quad (2.2)$$

where  $y_{k+1} = \frac{I_{k+1} - I_k}{I_k(N - I_k)\sqrt{\Delta t}}$ ,  $\eta = \mu + \gamma$ ,  $u_{k+1} = -\frac{\sqrt{\Delta t}}{N - I_k}$ ,  $c = \sqrt{\Delta t}\beta$  and  $Z_{k+1} \sim N(0, 1)$ . We can get the observations  $(y_i, u_i)_{i=1}^n$  if data points  $\{I_k\}_{k=0}^n$  and stepsize  $\Delta t$  are provided. We then write the model as  $y_i = \eta u_i + c + \varepsilon_i$  ( $i = 1, 2, \dots, n$ ), where  $\varepsilon_i \sim N(0, \sigma)$ . This looks like a simple linear regression model but with the difference that  $y = (y_1, y_2, \dots, y_n)$  is a random variable instead of a response variable which is conditional on  $u = (u_1, u_2, \dots, u_n)$ . However we still can use the regression theory to estimate  $\eta$  and  $\beta$  since estimation is based on the least squares method, i.e. to minimize  $\sum_{i=1}^n (y_i - \eta u_i - c)^2$ , which is not affected by whether  $y$  is a random variable or not.

Rawlings (1998) [22] discusses multiple linear regression in the general matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (2.3)$$

where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The calculations work equally well for (2.2), which can be written in the matrix form (2.3), where  $\mathbf{Y}$  and  $\boldsymbol{\varepsilon}$  remain the same while  $\mathbf{X}$  and  $\boldsymbol{\theta}$  become

$$\mathbf{X} = \begin{pmatrix} \sqrt{\Delta t} & u_1 \\ \sqrt{\Delta t} & u_2 \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_n \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \beta \\ \eta \end{pmatrix}.$$

## 2.2 Point Estimators

We use the formulae in the multiple linear regression theory to derive the estimators for  $\eta$  and  $\beta$  as

$$\begin{aligned} \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} &= \hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}) \\ &= \frac{1}{n\Delta t \sum u_k^2 - \Delta t (\sum u_k)^2} \begin{pmatrix} \sqrt{\Delta t} \sum u_k^2 \sum y_k - \sqrt{\Delta t} \sum u_k \sum u_k y_k \\ n\Delta t \sum u_k y_k - \Delta t \sum u_k \sum y_k \end{pmatrix}. \end{aligned} \quad (2.4)$$

Here  $\sum$  represents  $\sum_{k=0}^{n-1}$  as does the  $\sum$  below.

Then we have point estimators as

$$\hat{\beta} = \frac{\sum u_k^2 \sum y_k - \sum u_k \sum u_k y_k}{n\sqrt{\Delta t} \sum u_k^2 - \sqrt{\Delta t} (\sum u_k)^2} \quad (2.5)$$

and

$$\hat{\eta} = \frac{n \sum u_k y_k - \sum u_k \sum y_k}{n \sum u_k^2 - (\sum u_k)^2}, \quad (2.6)$$

which are equal to

$$\hat{\beta} = \frac{\sum \frac{1}{(N-I_k)^2} \sum \frac{I_{k+1}-I_k}{I_k(N-I_k)} - \sum \frac{1}{N-I_k} \sum \frac{I_{k+1}-I_k}{(N-I_k)^2 I_k}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \left( \sum \frac{\sqrt{\Delta t}}{N-I_k} \right)^2} \quad (2.7)$$

and

$$\hat{\eta} = \frac{\sum \frac{I_{k+1}-I_k}{I_k(N-I_k)} \sum \frac{1}{N-I_k} - n \sum \frac{I_{k+1}-I_k}{(N-I_k)^2 I_k}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \left( \sum \frac{\sqrt{\Delta t}}{N-I_k} \right)^2}. \quad (2.8)$$

We consider a time interval of total length  $T$  divided into  $n$  subintervals each of length  $\Delta t$  so  $n\Delta t = T$ . Hence as  $n \rightarrow \infty$  and  $\Delta t \rightarrow 0$  with  $n\Delta t = T$ , the sums approach the integrals, i.e.

$$\begin{aligned} \sum_{k=0}^{n-1} \frac{\Delta t}{(N-I_k)^2} &\rightarrow \int_0^T \frac{1}{(N-I)^2} dt \\ \sum_{k=0}^{n-1} \frac{I_{k+1}-I_k}{I_k(N-I_k)} &\rightarrow \int_{I(0)}^{I(T)} \frac{1}{I(N-I)} dI \quad \text{etc.} \end{aligned}$$

Hence we have that as  $n \rightarrow \infty$ ,  $\hat{\beta}$  and  $\hat{\eta}$  tend to

$$\hat{\beta} = \frac{\int_0^T \frac{1}{(N-I(t))^2} dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)I} dI - \int_0^T \frac{1}{N-I(t)} dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I} dI}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}$$

and

$$\hat{\eta} = \frac{\int_{I(0)}^{I(T)} \frac{1}{(N-I)I} dI \cdot \int_0^T \frac{1}{N-I(t)} dt - T \int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I} dI}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}.$$

## 2.3 Variance of Estimated Parameters

Confidence interval estimators of parameters give more information than simple point estimators. To obtain interval estimators for the parameters  $\beta$  and  $\eta$ , we need to calculate the variance of  $\hat{\boldsymbol{\theta}}$  using the formula

$$\text{var}(\hat{\boldsymbol{\theta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2, \quad (2.9)$$

where  $\sigma^2$  can be estimated using the residual mean square

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})}{n - p}, \quad (2.10)$$

where  $p$  is the number of parameters and is 2 in this case. Equation (2.10) can be simplified as

$$\hat{\sigma}^2 = \frac{\mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \hat{\boldsymbol{\theta}}}{n - 2} \quad (2.11)$$

if we substitute  $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y})$  in (2.10).

Then equation (2.11) can be written as

$$\hat{\sigma}^2 = \frac{1}{n - 2} \left( \sum y_k^2 - (\sqrt{\Delta t} \sum y_k) \hat{\beta} - \left( \sum y_k u_k \right) \hat{\eta} \right). \quad (2.12)$$

Substituting (2.4) in (2.12) we get

$\hat{\sigma}^2 =$

$$\frac{n \sum y_k^2 \sum u_k^2 - \sum y_k^2 (\sum u_k)^2 - \sum u_k^2 (\sum y_k)^2 - n (\sum y_k u_k)^2 + 2 \sum u_k \sum y_k \sum y_k u_k}{(n - 2) (n \sum u_k^2 - (\sum u_k)^2)}, \quad (2.13)$$

which is

$$\begin{aligned} & \frac{1}{n \left( n \sum \frac{\Delta t}{(N - I_k)^2} - \Delta t \left( \sum \frac{1}{N - I_k} \right)^2 \right)} \\ & \left( n \sum \frac{(I_{k+1} - I_k)^2}{I_k^2 (N - I_k)^2} \sum \frac{1}{(N - I_k)^2} - \sum \frac{(I_{k+1} - I_k)^2}{I_k^2 (N - I_k)^2} \left( \sum \frac{1}{N - I_k} \right)^2 \right. \\ & - \sum \frac{1}{(N - I_k)^2} \left( \sum \frac{I_{k+1} - I_k}{I_k (N - I_k)} \right)^2 - n \left( \sum \frac{I_{k+1} - I_k}{I_k (N - I_k)^2} \right)^2 \\ & \left. + 2 \sum \frac{1}{N - I_k} \sum \frac{I_{k+1} - I_k}{I_k (N - I_k)} \sum \frac{I_{k+1} - I_k}{I_k (N - I_k)^2} \right). \end{aligned} \quad (2.14)$$

**Theorem 2.1** *The estimator  $\hat{\sigma}^2$  in (2.12) is an asymptotically unbiased estimator for  $\sigma^2$  in (2.2), i.e.*

$$\hat{\sigma}_n^2 \rightarrow \sigma^2 \quad \text{a.s.}$$

as  $n \rightarrow \infty$ .

*Proof.*

$$\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{1}{n-2} \sum (y_k - \hat{y}_k)^2, \quad (2.15)$$

where RSS is the sum of squares of residuals for model (2.2) and  $p$  is the number of parameters to be estimated.

After substituting for  $\hat{\beta}$  and  $\hat{\eta}$  using (2.5) and (2.6)

$$\begin{aligned} y_k - \hat{y}_k &= y_k \\ &\quad - \frac{\sum_i u_i^2 \sum_i y_i - \sum_i u_i \sum_i u_i y_i}{\Delta} \\ &\quad - \frac{(n \sum_i u_i y_i - \sum_i u_i \sum_i y_i) u_k}{\Delta}, \end{aligned}$$

where  $\Delta = n \sum_i u_i^2 - (\sum_i u_i)^2$  and  $\sum_i$  represents  $\sum_{i=0}^{n-1}$  here and throughout the rest of the paper.

Since  $y_k = \beta\sqrt{\Delta t} + \eta u_k + \sigma Z_k$ ,

$$\begin{aligned} y_k - \hat{y}_k &= \beta\sqrt{\Delta t} + \eta u_k + \sigma Z_k \\ &\quad - \frac{\sum_i u_i^2 \sum_i (\beta\sqrt{\Delta t} + \eta u_i + \sigma Z_i) - \sum_i u_i \sum_i u_i (\beta\sqrt{\Delta t} + \eta u_i + \sigma Z_i)}{\Delta} \\ &\quad - \frac{[n \sum_i u_i (\beta\sqrt{\Delta t} + \eta u_i + \sigma Z_i) - \sum_i u_i \sum_i (\beta\sqrt{\Delta t} + \eta u_i + \sigma Z_i)] u_k}{\Delta}. \end{aligned}$$

Therefore  $\hat{\sigma}^2$  can be simplified as

$$\frac{\sigma^2}{n-2} \sum \left( Z_k - \frac{u_k n \sum_i u_i Z_i - u_k \sum_i Z_i \sum_i u_i + \sum_i u_i^2 \sum_i Z_i - \sum_i u_i \sum_i u_i Z_i}{\Delta} \right)^2,$$

which is equal to

$$\begin{aligned} \frac{\sigma^2}{n-2} \sum &\left( Z_k^2 + \left( \frac{n u_k \sum_i u_i Z_i - u_k \sum_i Z_i \sum_i u_i}{\Delta} \right)^2 + \left( \frac{\sum_i u_i^2 \sum_i Z_i - \sum_i u_i \sum_i u_i Z_i}{\Delta} \right)^2 \right. \\ &\quad - 2 Z_k \frac{n u_k \sum_i u_i Z_i - u_k \sum_i Z_i \sum_i u_i}{\Delta} - 2 Z_k \frac{\sum_i u_i^2 \sum_i Z_i - \sum_i u_i \sum_i u_i Z_i}{\Delta} \\ &\quad \left. + 2 \frac{(u_k n \sum_i u_i Z_i - u_k \sum_i Z_i \sum_i u_i) (\sum_i u_i^2 \sum_i Z_i - \sum_i u_i \sum_i u_i Z_i)}{\Delta^2} \right). \end{aligned}$$

This can be simplified as

$$\begin{aligned} \frac{\sigma^2}{n-2} &\left( \sum Z_k^2 + \frac{1}{\Delta^2} \left( \sum u_k^2 \left( \sum Z_k \right)^2 \left( \sum u_k \right)^2 - n^2 \left( \sum u_k Z_k \right)^2 \sum u_k^2 \right. \right. \\ &\quad - n \left( \sum u_k^2 \right)^2 \left( \sum Z_k \right)^2 + n \left( \sum u_k Z_k \right)^2 \left( \sum u_k \right)^2 \\ &\quad \left. \left. - 2 \left( \sum u_k \right)^3 \sum Z_k \sum u_k Z_k + 2n \sum u_k \sum u_k Z_k \sum u_k^2 \sum Z_k \right) \right), \end{aligned}$$

which equals

$$\frac{\sigma^2}{n-2} \left( \sum Z_k^2 + \frac{-n(\sum u_k Z_k)^2 - \sum u_k^2 (\sum Z_k)^2 + 2 \sum u_k \sum u_k Z_k \sum Z_k}{n \sum u_k^2 - (\sum u_k)^2} \right).$$

This can be written as

$$\frac{\sigma^2}{n-2} \left( \sum Z_k^2 - A \cdot \frac{(\sum u_k Z_k)^2}{\sum u_k^2} - A \cdot \frac{(\sum Z_k)^2}{n} + B \cdot \frac{\sum u_k Z_k}{\sqrt{\sum u_k^2}} \cdot \frac{\sum Z_k}{\sqrt{n}} \right), \quad (2.16)$$

where

$$A = \frac{n \sum u_k^2}{n \sum u_k^2 - (\sum u_k)^2}, \quad B = 2 \cdot \frac{\sqrt{n} \sqrt{\sum u_k^2} \sum u_k}{n \sum u_k^2 - (\sum u_k)^2}.$$

Note that

$$\frac{\sum Z_k}{\sqrt{n}} \sim N(0, 1) \quad \text{and} \quad \frac{\sum u_k Z_k}{\sqrt{\sum u_k^2}} \sim N(0, 1),$$

since  $Z_k \sim N(0, 1)$  and the  $Z_k$  are independent. Moreover  $\frac{(\sum Z_k)^2}{n}$  and  $\frac{(\sum u_k Z_k)^2}{\sum u_k^2}$  have mean 1 and variance  $v$ , where  $v$  can be worked out and is independent of  $n$  (it is the variance of the square of a standard normal random variable).

Therefore

$$\frac{\sigma^2}{n-2} \left( -A \cdot \frac{(\sum u_k Z_k)^2}{\sum u_k^2} - A \cdot \frac{(\sum Z_k)^2}{n} + B \cdot \frac{\sum u_k Z_k}{\sqrt{\sum u_k^2}} \cdot \frac{\sum Z_k}{\sqrt{n}} \right) \rightarrow 0 \quad a.s.$$

as  $n \rightarrow \infty$ .

Also  $Z_k^2$  has mean 1 and variance  $v_2$  independent of  $n$  since  $Z_k \sim N(0, 1)$ . Therefore  $\frac{1}{n} \sum Z_k^2 \rightarrow N(1, \frac{v_2}{n})$  as  $n \rightarrow \infty$  by the Central Limit Theorem.

Hence  $\hat{\sigma}^2 \rightarrow \sigma^2$  with probability one as  $n \rightarrow \infty$  as required.  $\square$

Using  $\hat{\sigma}^2$  to estimate  $\sigma^2$  in (2.9) we have

$$\text{var}(\hat{\boldsymbol{\theta}}) = \text{var} \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} = \frac{1}{n\Delta t \sum u_k^2 - \Delta t (\sum u_k)^2} \begin{pmatrix} \sum u_k^2 & -\sqrt{\Delta t} \sum u_k \\ -\sqrt{\Delta t} \sum u_k & n\Delta t \end{pmatrix} \hat{\sigma}^2. \quad (2.17)$$

## 2.4 Interval Estimation

The distribution of the parameter estimators  $\hat{\beta}$  and  $\hat{\eta}$  if  $\sigma^2$  is known is exactly multivariate (actually bivariate) normal by the least squares regression theory [22]. However we are estimating  $\sigma^2$  by  $\hat{\sigma}^2$ . Hence if the number of observations  $n$  is large, the approximate  $100(1 - \alpha)\%$  confidence intervals (CIs) for  $\beta$  and  $\eta$  respectively are

$$\hat{\beta} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\beta})} = \frac{\sum u_k^2 \sum y_k - \sum u_k \sum u_k y_k}{n\sqrt{\Delta t} \sum u_k^2 - \sqrt{\Delta t} (\sum u_k)^2} \pm z_{\alpha/2} \sqrt{\frac{\sum u_k^2 \hat{\sigma}^2}{n\Delta t \sum u_k^2 - \Delta t (\sum u_k)^2}} \quad (2.18)$$

and

$$\hat{\eta} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\eta})} = \frac{n \sum u_k y_k - \sum u_k \sum y_k}{n \sum u_k^2 - (\sum u_k)^2} \pm z_{\alpha/2} \sqrt{\frac{n \Delta t \hat{\sigma}^2}{n \Delta t \sum u_k^2 - \Delta t (\sum u_k)^2}}, \quad (2.19)$$

where  $z_{\alpha/2}$  is the upper  $\alpha/2$  value of the standard normal random variable, e.g.  $z_{0.025} = 1.96$  for a 95% CI.

We notice that as  $n \rightarrow \infty$ , the  $100(1 - \alpha)\%$  CIs tend to

$$\frac{\int_0^T \frac{1}{(N-I(t))^2} dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)I} dI - \int_0^T \frac{1}{N-I(t)} dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I} dI}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2} \pm z_{\alpha/2} \sqrt{\frac{\int_0^T \frac{1}{(N-I(t))^2} dt \cdot \sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}}$$

and

$$\frac{\int_{I(0)}^{I(T)} \frac{1}{(N-I)I} dI \cdot \int_0^T \frac{1}{N-I(t)} dt - T \int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I} dI}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2} \pm z_{\alpha/2} \sqrt{\frac{T \sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}},$$

respectively.

**Theorem 2.2** *The asymptotic widths of the CIs for both  $\beta$  and  $\eta$ , which are*

$$2 \times z_{\alpha/2} \sqrt{\frac{\int_0^T \frac{1}{(N-I(t))^2} dt \cdot \sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}}$$

and

$$2 \times z_{\alpha/2} \sqrt{\frac{T \sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}},$$

are strictly decreasing as  $T$  increases.

*Proof.* Considering first the width of the CI for  $\beta$ ,

$$\frac{\int_0^T \frac{1}{(N-I(t))^2} dt}{T \int_0^T \frac{1}{(N-I(t))^2} dt - \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2} = \frac{1}{T - \frac{\left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}{\int_0^T \frac{1}{(N-I(t))^2} dt}}. \quad (2.20)$$



Then the derivative of the denominator is equal to

$$\begin{aligned}
& \frac{d}{dT} \left( T - \frac{\left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}{\int_0^T \frac{1}{(N-I(t))^2} dt} \right) \\
&= 1 - \frac{2 \int_0^T \frac{1}{N-I(t)} dt \frac{1}{N-I(T)} \int_0^T \frac{1}{(N-I(t))^2} dt - \frac{1}{(N-I(T))^2} \left( \int_0^T \frac{1}{N-I(t)} dt \right)^2}{\left( \int_0^T \frac{1}{(N-I(t))^2} dt \right)^2} \\
&= \frac{\left( \int_0^T \frac{1}{(N-I(t))^2} dt - \frac{1}{N-I(T)} \int_0^T \frac{1}{N-I(t)} dt \right)^2}{\left( \int_0^T \frac{1}{(N-I(t))^2} dt \right)^2} \geq 0.
\end{aligned}$$

Given a sample path  $I(t)$  defined on the interval  $[0, T]$  with  $I(0) > 0$ , we deduce that  $I(T) \in (0, N)$  [6]. The only way that the denominator of (2.20) is not strictly decreasing is if

$$\int_0^T \frac{1}{(N-I(t))^2} dt = \frac{1}{N-I(T)} \int_0^T \frac{1}{N-I(t)} dt$$

on an interval  $[T, T + \varepsilon]$  for some  $\varepsilon > 0$ .

So if  $\Delta T$  is small enough

$$(N - I(T + \Delta T)) \int_0^{T+\Delta T} \frac{1}{(N-I(t))^2} dt = \int_0^{T+\Delta T} \frac{1}{N-I(t)} dt \quad (2.21)$$

and

$$(N - I(T)) \int_0^T \frac{1}{(N-I(t))^2} dt = \int_0^T \frac{1}{N-I(t)} dt. \quad (2.22)$$

Subtracting (2.22) from (2.21) we have

$$\begin{aligned}
& [(N - I(T + \Delta T)) - (N - I(T))] \int_0^{T+\Delta T} \frac{1}{(N-I(t))^2} dt \\
& + (N - I(T)) \left( \int_0^{T+\Delta T} \frac{1}{(N-I(t))^2} dt - \int_0^T \frac{1}{(N-I(t))^2} dt \right) \\
& = \int_0^{T+\Delta T} \frac{1}{N-I(t)} dt - \int_0^T \frac{1}{N-I(t)} dt,
\end{aligned}$$

which is equal to

$$\begin{aligned}
& (-I(T + \Delta T) + I(T)) \left( \int_0^T \frac{1}{(N-I(t))^2} dt + \frac{\Delta T}{(N-I(T))^2} + o(\Delta T) \right) \\
& + (N - I(T)) \left( \frac{\Delta T}{(N-I(T))^2} + o(\Delta T) \right) = \frac{\Delta T}{N-I(T)} + o(\Delta T).
\end{aligned}$$

This equals

$$-I(T) \left( (\beta N - \eta - \beta I(T)) \Delta T + \sigma(N - I(T)) [B(T + \Delta T) - B(T)] + o(\Delta T) \right) \\ \int_0^T \frac{1}{(N - I(t))^2} dt = o(\Delta T).$$

Dividing by  $\sqrt{\Delta T}$  we have

$$-I(T) \left( (\beta N - \eta - \beta I(T)) \sqrt{\Delta T} + \sigma(N - I(T)) \frac{B(T + \Delta T) - B(T)}{\sqrt{\Delta T}} \right) = o(\sqrt{\Delta T}).$$

Letting the time step  $\Delta T$  be very small and choosing  $\varepsilon_0 > 0$ ,  $\exists \Delta T_0 \leq 1$  such that for  $\Delta T < \Delta T_0$  the  $o(\sqrt{\Delta T})$  term is between  $-\varepsilon_0 \sigma I(T) (N - I(T)) \sqrt{\Delta T}$  and  $+\varepsilon_0 \sigma I(T) (N - I(T)) \sqrt{\Delta T}$ , hence must lie between  $-\varepsilon_0 \sigma I(T) (N - I(T))$  and  $\varepsilon_0 \sigma I(T) (N - I(T))$ .

Hence the term

$$\frac{B(T + \Delta T) - B(T)}{\sqrt{\Delta T}} \in \left( -\frac{\beta N - \eta - \beta I(T)}{\sigma(N - I(T))} - \varepsilon_0, -\frac{\beta N - \eta - \beta I(T)}{\sigma(N - I(T))} + \varepsilon_0 \right).$$

But  $\frac{B(T + \Delta T) - B(T)}{\sqrt{\Delta T}} \sim N(0, 1)$  so the probability that it lies in the above interval tends to zero as  $\varepsilon_0 \rightarrow 0$ . Hence

$$P \left( \int_0^T \frac{1}{(N - I(t))^2} dt = \frac{1}{N - I(T)} \int \frac{1}{N - I(t)} dt \text{ on } [T, T + \varepsilon] \text{ for some } \varepsilon > 0 \right) = 0.$$

So the denominator of (2.20) is strictly increasing and the width of the CI for  $\beta$  is strictly decreasing in  $T$ . Similarly we can prove the case for  $\eta$ .  $\square$

In the same way as for the simple linear SDE case, the asymptotic widths of the CIs here do not depend on the size of time step  $\Delta t$  but only on the total time period  $T$ , and are decreasing as  $T$  increases.

## 2.5 Joint Confidence Region

We have obtained univariate CIs for each parameter  $\beta$  and  $\eta$  in the last section. However individual CIs do not take into account the correlation among the parameters. Also, they do not reflect the overall degree of confidence. Joint confidence regions take both issues into account. So we will obtain a joint confidence region for  $\beta$  and  $\eta$  in this section.

A  $100(1 - \alpha)\%$  joint confidence region for the general regression model (2.3) is obtained from the following inequality [22]

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (\mathbf{X}^T \mathbf{X}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq p \sigma^2 F_{\alpha, p, \nu} \quad (2.23)$$

where  $F_{\alpha, p, \nu}$  is the value of the  $F$ -distribution with degrees of freedom  $p$  and  $\nu$  that leaves probability  $\alpha$  in the upper tail,  $p$  is the number of parameters and  $\nu$  is the degrees of freedom associated with  $\sigma^2$ .

Our case only involves two parameters so the  $100(1 - \alpha)\%$  joint confidence region for  $\beta$  and  $\eta$  can be written as

$$\left( \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} - \begin{pmatrix} \beta \\ \eta \end{pmatrix} \right)^T (\text{var}(\hat{\beta}, \hat{\eta}))^{-1} \left( \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} - \begin{pmatrix} \beta \\ \eta \end{pmatrix} \right) \leq 2F_{\alpha, 2, n-2}. \quad (2.24)$$

After substituting (2.17) in (2.24), it can be easily calculated as

$$n\Delta t (\hat{\beta} - \beta)^2 + 2\sqrt{\Delta t} \sum u_k (\hat{\beta} - \beta)(\hat{\eta} - \eta) + \sum u_k^2 (\hat{\eta} - \eta)^2 \leq 2\hat{\sigma}^2 F_{\alpha, 2, n-2}, \quad (2.25)$$

where  $\hat{\beta}$  and  $\hat{\eta}$  are known in (2.4).

We compute that

$$4\Delta t \left( \sum u_k \right)^2 - 4n\Delta t \sum u_k^2 = 4\Delta t \left( \left( \sum u_k \right)^2 - n \sum u_k^2 \right). \quad (2.26)$$

Defining the vectors in  $\mathbb{R}^n$

$$\mathbf{a} = (1, 1, \dots, 1) \quad \text{and} \quad \mathbf{b} = (u_1, u_2, \dots, u_n).$$

then  $|\mathbf{a}|^2|\mathbf{b}|^2 \geq |\mathbf{a} \cdot \mathbf{b}|^2$  as  $|\mathbf{a} \cdot \mathbf{b}| = |\mathbf{a}||\mathbf{b}|\cos\theta$ , where  $\theta$  is the angle between  $\mathbf{a}$  and  $\mathbf{b}$ . Then we have  $|\mathbf{a}|^2 = n$ ,  $|\mathbf{b}|^2 = \sum u_k^2$  and  $|\mathbf{a} \cdot \mathbf{b}| = (\sum u_k)^2$ . So (2.26) is strictly negative since the angle between  $\mathbf{a}$  and  $\mathbf{b}$  is not 0. Therefore the boundary of the  $100(1 - \alpha)\%$  joint confidence region is an ellipsoid.

**Example 2.5.1** Assume that the parameters are given by  $T = 1$ ,  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$ , and  $\sigma^2 = 0.03$  for the model (1.2).  $T = 1$  is scaled to represent one year and  $N = 100$  represents 100 million people. We will use the same scaling for the rest of the examples.

We simulate  $I(t)$  using the above parameters by the EM method with a very small step size  $\Delta t = 0.001$  and save these  $I(t)$  as our true data set. Then we sample every 10th data point in the data set to obtain the sample for our parameter estimation, so  $n = 100$  observations and  $\Delta t = 0.01$  for our sample.

With the sample we obtained, with  $\alpha = 0.05$  we find the boundary of the 95% joint confidence region for  $\beta$  and  $\eta$  by using (2.25) and the univariate 95% CIs for each of them by using (2.18) and (2.19), and also the point estimators by using (2.5) and (2.6), which are shown in Figure 1a. The ellipsoid in the figure represents the 95% joint confidence region while the grey lines represent the univariate CIs. We see that most of the ellipsoid lies in the square which represents the univariate CIs, but the area of the ellipsoid is much smaller than the square. It indicates the advantage of the joint confidence region that it drops out many of the extreme values in the univariate CIs and is more efficient. We see that the ellipsoid is centered at the point estimators of  $\beta$  and  $\eta$ . The red point which represents the true value of  $\beta = 0.5$  and  $\eta = 45$  lies in the ellipse.

**Example 2.5.2** Assume that the parameters are given by  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$ , and  $\sigma^2 = 0.03$  for the model (1.2), as in Example 2.5.1.

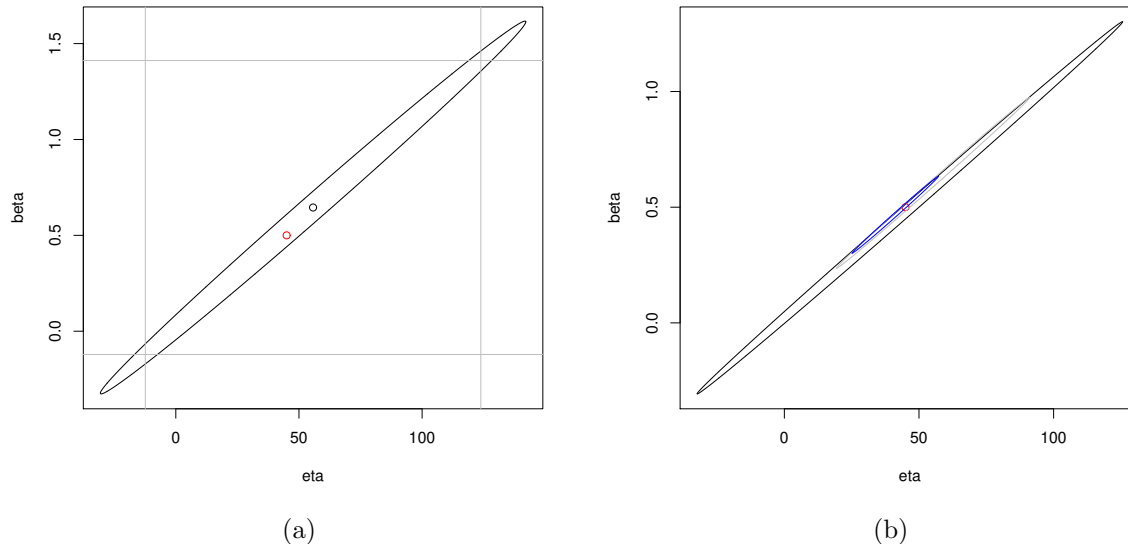


Figure 1: (a) is the 95% joint confidence region for  $\beta$  and  $\eta$  obtained using the parameter values in Example 2.5.1 with  $T = 1$ . The ellipsoid in the figure represents the 95% joint confidence region, while the grey vertical and horizontal lines represent the univariate CIs for each of  $\beta$  and  $\eta$ . The black point marked in the figure is the point estimate for  $\beta$  and  $\eta$ , and the red point represents the true values of  $\beta = 0.5$  and  $\eta = 45$ ; (b) is the 95% joint confidence region for  $\beta$  and  $\eta$  using the parameter values in Example 2.5.2, with  $T = 5$  (black),  $T = 20$  (grey) and  $T = 50$  (blue). The red point represents the true value of  $\beta = 0.5$  and  $\eta = 45$ .

*In order to see the influence of different interval lengths  $T$  on the 95% joint confidence region, we now vary the value of the interval length  $T$  as  $T = 5$ ,  $T = 20$  and  $T = 50$  and use the same method as in Example 2.5.1 to simulate a data set for each  $T$  and sample from each of them. When we increase  $T$  we increase the number of observations  $n$  in proportion to  $T$  to keep  $\Delta t$  fixed. We then obtain the three 95% joint confidence regions for the different values of  $T$ , which are shown in Figure 1b. We see that the area of the 95% joint confidence region becomes smaller with larger  $T$  (larger sequence of observations  $n$ ). The red point which represents the true value of  $\beta = 0.5$  and  $\eta = 45$  lies in all the ellipses.*

## 2.6 Estimation from Improved Regression Model with More Data Sets

The CIs for both  $\beta$  and  $\eta$  are dependent on the sample path. If more data sets are available, we can expand the original regression model to get better parameter estimation.

Assuming that we have  $m$  data sets each of size  $n$ , we can put all these data sets in

the regression model (2.2), so that  $\mathbf{Y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\theta}$  and  $\boldsymbol{\varepsilon}$  become

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \\ \vdots \\ y_{m1} \\ y_{m2} \\ \vdots \\ y_{mn} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \sqrt{\Delta t} & u_{11} \\ \sqrt{\Delta t} & u_{12} \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_{1n} \\ \sqrt{\Delta t} & u_{21} \\ \sqrt{\Delta t} & u_{22} \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_{2n} \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_{m1} \\ \sqrt{\Delta t} & u_{m2} \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_{mn} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \beta \\ \eta \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \\ \vdots \\ \varepsilon_{m1} \\ \varepsilon_{m2} \\ \vdots \\ \varepsilon_{mn} \end{pmatrix},$$

using the same formula as in (2.4) we have

$$\begin{aligned} \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} &= \hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}) \\ &= \frac{1}{mn\Delta t \sum \sum u_{ij}^2 - \Delta t (\sum \sum u_{ij})^2} \begin{pmatrix} \sqrt{\Delta t} \sum \sum u_{ij}^2 \sum \sum y_{ij} - \sqrt{\Delta t} \sum \sum u_{ij} \sum \sum u_{ij} y_{ij} \\ mn\Delta t \sum \sum u_{ij} y_{ij} - \Delta t \sum \sum u_{ij} \sum \sum y_{ij} \end{pmatrix}, \end{aligned} \quad (2.27)$$

where  $\sum \sum = \sum_{i=1}^m \sum_{j=0}^{n-1}$  and similarly below.

In the same way, we can get

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \hat{\boldsymbol{\theta}}}{mn - 2} \\ &= \frac{1}{mn - 2} \left( \sum \sum y_{ij}^2 - (\sqrt{\Delta t} \sum \sum y_{ij}) \hat{\beta} - (\sum \sum y_{ij} u_{ij}) \hat{\eta} \right) \\ &= \frac{1}{(mn - 2)(mn\Delta t \sum \sum u_{ij}^2 - \Delta t (\sum \sum u_{ij})^2)} \cdot \left( mn\Delta t \sum \sum y_{ij}^2 \sum \sum u_{ij}^2 \right. \\ &\quad - \Delta t \sum \sum y_{ij}^2 (\sum \sum u_{ij})^2 - \Delta t \sum \sum u_{ij}^2 (\sum \sum y_{ij})^2 \\ &\quad \left. - mn\Delta t (\sum \sum y_{ij} u_{ij})^2 + 2\Delta t \sum \sum u_{ij} \sum \sum y_{ij} \sum \sum y_{ij} u_{ij} \right). \end{aligned} \quad (2.28)$$

When proving  $\hat{\sigma}^2$  is an asymptotically unbiased estimator of  $\sigma^2$ , the procedure is similar to the one we used before. We use an equation similar to (2.15),

$$\hat{\sigma}^2 = \frac{1}{mn - 2} \sum \sum (y_{ij} - \hat{y}_{ij})^2.$$

After almost identical working to that used before, we can simplify  $\hat{\sigma}^2$  as in (2.16), except that the  $\sum$  now represents  $\sum_{i=1}^m \sum_{j=0}^{n-1}$  and the denominator under  $\sigma^2$  is  $mn - 2$ . We know that  $n \rightarrow \infty$  implies  $mn \rightarrow \infty$ .

So following almost the same procedure for the proof as before, we can prove that  $\hat{\sigma}^2 \rightarrow \sigma^2$  with probability one as  $n \rightarrow \infty$ .

Using formula (2.9) and  $\hat{\sigma}$  in (2.28) to estimate  $\sigma$  we have

$$\text{var}(\hat{\theta}) = \text{var} \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} = \frac{1}{mn\Delta t \sum \sum u_{ij}^2 - \Delta t (\sum \sum u_{ij})^2} \begin{pmatrix} \sum \sum u_{ij}^2 - \sqrt{\Delta t} \sum \sum u_{ij} \\ -\sqrt{\Delta t} \sum \sum u_{ij} mn\Delta t \end{pmatrix} \hat{\sigma}^2. \quad (2.29)$$

If the number of observations is large, the  $100(1 - \alpha)\%$  CIs for  $\beta$  and  $\eta$  estimated from the regression model with  $m$  data sets are

$$\begin{aligned} & \hat{\beta} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\beta})} \\ &= \frac{\sum \sum u_{ij}^2 \sum \sum y_{ij} - \sum \sum u_{ij} \sum \sum u_{ij} y_{ij}}{mn\sqrt{\Delta t} \sum \sum u_{ij}^2 - \sqrt{\Delta t} (\sum \sum u_{ij})^2} \pm z_{\alpha/2} \sqrt{\frac{\sum \sum u_{ij}^2 \hat{\sigma}^2}{mn\Delta t \sum \sum u_{ij}^2 - \Delta t (\sum \sum u_{ij})^2}} \end{aligned}$$

and

$$\begin{aligned} & \hat{\eta} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\eta})} \\ &= \frac{mn \sum \sum u_{ij} y_{ij} - \sum \sum u_{ij} \sum \sum y_{ij}}{mn \sum \sum u_{ij}^2 - (\sum \sum u_{ij})^2} \pm z_{\alpha/2} \sqrt{\frac{mn\Delta t \hat{\sigma}^2}{mn\Delta t \sum \sum u_{ij}^2 - \Delta t (\sum \sum u_{ij})^2}}, \end{aligned}$$

respectively.

As  $n \rightarrow \infty$ , the  $100(1 - \alpha)\%$  CIs tend to

$$\begin{aligned} & \frac{\sum \int_0^T \frac{1}{(N-I_i(t))^2} dt \cdot \sum \int_{I_i(0)}^{I_i(T)} \frac{1}{(N-I_i)I_i} dI_i - \sum \int_0^T \frac{1}{N-I_i(t)} dt \cdot \sum \int_{I_i(0)}^{I_i(T)} \frac{1}{(N-I_i)^2 I_i} dI_i}{mT \sum \int_0^T \frac{1}{(N-I_i(t))^2} dt - (\sum \int_0^T \frac{1}{N-I_i(t)} dt)^2} \\ & \pm z_{\alpha/2} \sqrt{\frac{\sum \int_0^T \frac{1}{(N-I_i(t))^2} dt \cdot \sigma^2}{mT \sum \int_0^T \frac{1}{(N-I_i(t))^2} dt - (\sum \int_0^T \frac{1}{N-I_i(t)} dt)^2}} \end{aligned} \quad (2.30)$$

and

$$\begin{aligned} & \frac{\sum \int_{I_i(0)}^{I_i(T)} \frac{1}{(N-I_i)I_i} dI_i \cdot \sum \int_0^T \frac{1}{N-I_i(t)} dt - mT \sum \int_{I_i(0)}^{I_i(T)} \frac{1}{(N-I_i)^2 I_i} dI_i}{mT \sum \int_0^T \frac{1}{(N-I_i(t))^2} dt - (\sum \int_0^T \frac{1}{N-I_i(t)} dt)^2} \\ & \pm z_{\alpha/2} \sqrt{\frac{mT \sigma^2}{mT \sum \int_0^T \frac{1}{(N-I_i(t))^2} dt - (\sum \int_0^T \frac{1}{N-I_i(t)} dt)^2}}, \end{aligned} \quad (2.31)$$

respectively. Here  $\sum$  represents  $\sum_{i=1}^m$ .

**Example 2.6.1** Assume that the parameters are given by  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$ , and  $\sigma^2 = 0.03$  for the model (1.2), as previously.

In this example we compare the following 3 methods in terms of the efficiency of interval estimation. Method 1: One observer is assigned to record  $I(t)$  at one location four times more densely than the comparison during  $T$ . Method 2: Two observers are assigned to record  $I(t)$  with the same time steps as the comparison at four locations during  $T$  and these four samples are combined for estimation. Method 3: one observer is assigned to record  $I(t)$  with the same time steps as the comparison during time period  $4T$ . To achieve this purpose we design the experiment as follows:

We obtain three data sets as in Example 2.5.1, 5 times. The first 4 data sets use the model parameters above and  $T = 25$ , while the fifth data set uses  $T = 100$ . Then we sample every 20th data point in the first data set to obtain sample A, so  $n = 1250$  and  $\Delta t = 0.02$  for this case. We use sample A as the benchmark. For Method 1, we obtain sample B by sampling every 5th data point in the first data set, so  $n = 5000$  and  $\Delta t = 0.005$  for this case. We then use (2.18) and (2.19) to obtain the 95% CIs for  $\beta$  and  $\eta$ . For method 2, we sample every 20th data point in the 2nd to 4th data sets to get sample C, D, E and combine them with sample A to obtain 4 samples each of  $n = 1250$  and  $\Delta t = 0.02$ . For sample A, C, D and E combined together we have  $n = 5000$  observations in total and  $\Delta t = 0.02$ . We then use estimators from the regression model with more data sets using (2.30) and (2.31) to obtain the 95% CIs for  $\beta$  and  $\eta$  (with  $\alpha = 0.05$ ,  $z_{\alpha/2} = 1.96$ ). For method 3, we sample every 20th data point in the 5th data set to obtain sample F so  $n = 5000$  and  $\Delta t = 0.02$  for this case. The results are displayed in Table 1.

We see from Table 1 that Method 1 (Sample B), using a sample from one location with denser observations, does not give smaller CIs for both  $\beta$  and  $\eta$ , while Method 2 (Samples A, C, D and E), using more samples at different locations, decreases the width of the CIs significantly and improves the efficiency of estimation. Method 3 (Sample F), using a sample with longer observations at one location, also gives narrower CIs. Therefore we conclude from this example that both Method 2 and 3 improve the efficiency of estimation. We have repeated our simulations with different model parameter values and the conclusions are the same.

Table 1: CIs for Example 2.6.1; results are repeated 3 times.

	CI for $\beta$	Width of CI	CI for $\eta$	Width of CI
Sample A ( $n = 1250$ , $\Delta t = 0.02$ )	(0.36, 1.12)	0.76	(32.06, 106.58)	74.52
	(0.32, 0.78)	0.46	(24.78, 71.42)	42.64
	(0.15, 0.60)	0.45	(10.84, 54.63)	43.79
Sample B ( $n = 5000$ , $\Delta t = 0.005$ )	(0.41, 1.13)	0.72	(36.77, 107.53)	70.76
	(0.31, 0.72)	0.41	(26.16, 65.84)	39.68
	(0.19, 0.62)	0.43	(14.35, 56.29)	41.94
Sample A, C, D and E ( $4 \times n = 1250$ , $\Delta t = 0.02$ )	(0.36, 0.68)	0.32	(30.98, 62.92)	31.94
	(0.39, 0.62)	0.23	(33.68, 56.15)	22.47
	(0.38, 0.59)	0.22	(32.62, 53.50)	20.88
Sample F ( $n = 5000$ , $\Delta t = 0.02$ )	(0.43, 0.76)	0.33	(37.72, 71.73)	34.01
	(0.33, 0.63)	0.30	(28.28, 58.07)	29.79
	(0.38, 0.60)	0.22	(33.12, 54.34)	21.22

By substituting (2.29) in (2.24), we can easily work out the joint  $100(1 - \alpha)\%$  confi-

dence region for  $\beta$  and  $\eta$  for the regression model with  $m$  data sets of size  $n$  as

$$mn\Delta t(\hat{\beta}-\beta)^2+2\sqrt{\Delta t}\sum\sum u_{ij}(\hat{\beta}-\beta)(\hat{\eta}-\eta)+\sum\sum u_{ij}^2(\hat{\eta}-\eta)^2\leq 2\delta^2F_{\alpha,2,mn-2}, \quad (2.32)$$

where  $\hat{\beta}$  and  $\hat{\eta}$  are given in (2.27).

We compute that

$$4\Delta t\left(\sum\sum u_{ij}\right)^2-4mn\Delta t\sum\sum u_{ij}^2=4\Delta t\left(\left(\sum\sum u_{ij}\right)^2-mn\sum\sum u_{ij}^2\right). \quad (2.33)$$

As for the regression model with one data set in section 2.5, we can prove that (2.33) is strictly negative. Therefore the boundary of the  $100(1-\alpha)\%$  joint confidence region for the regression model with  $m$  data sets is still an ellipsoid.

**Example 2.6.2** Assume that the parameters are given by  $T = 1$ ,  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$ ,  $m = 10$  and  $\sigma^2 = 0.03$  for the model (1.2).

We simulate  $I(t)$  using the above parameters by the EM method with a very small step size  $\Delta t = 0.001$ ,  $m = 10$  times and save these  $I(t)$  as 10 sets of true data. Then we sample every 10th data in each data set to obtain 10 samples for our parameter estimation, so  $n = 100$  and  $\Delta t = 0.01$  for each of our samples.

We find the boundary of the 95% joint confidence region for  $\beta$  and  $\eta$  using (2.33) and the univariate 95% CIs for each of them using (2.30) and (2.31), and also the point estimates using (2.27). These are shown in Figure 2(a).

We see that most of the ellipse lies in the square which represents the univariate CIs, but the area of the ellipse is much smaller than that of the square. This indicates the advantage of the joint confidence region, i.e. it does not include many of the extreme values in the univariate CIs and is more efficient. Also we see that the ellipse is centered at the point estimate of  $\beta$  and  $\eta$ . The red point which represents the true value of  $\beta = 0.5$  and  $\eta = 45$  lies in the ellipse.

**Example 2.6.3** Assume that the model parameters are given by  $T = 1$ ,  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$  and  $\sigma^2 = 0.03$  for the model (1.2).

In order to examine the influence of different  $m$  on the 95% joint confidence region, we vary the value of  $m$  as  $m = 1$ ,  $m = 2$  and  $m = 5$  and use the same method as in Example 2.6.2 to simulate data sets for each  $m$  and sample from each of them. We then obtain three 95% joint confidence regions for the different  $m$ , which are shown in Figure 2(b). We see that the area of the 95% joint confidence region becomes smaller as  $m$  becomes larger. Also the red point which represents the true value of  $\beta = 0.5$  and  $\eta = 45$  lies in all the ellipses.

### 3 Pseudo-Maximum Likelihood Estimation

In this context, the explicit expressions for MLEs for  $\phi = (\beta, \eta, \sigma^2)$  are not attainable, primarily because it is very difficult to find the corresponding likelihood function. Therefore we are unable to use the exact Maximum Likelihood method. An approximation



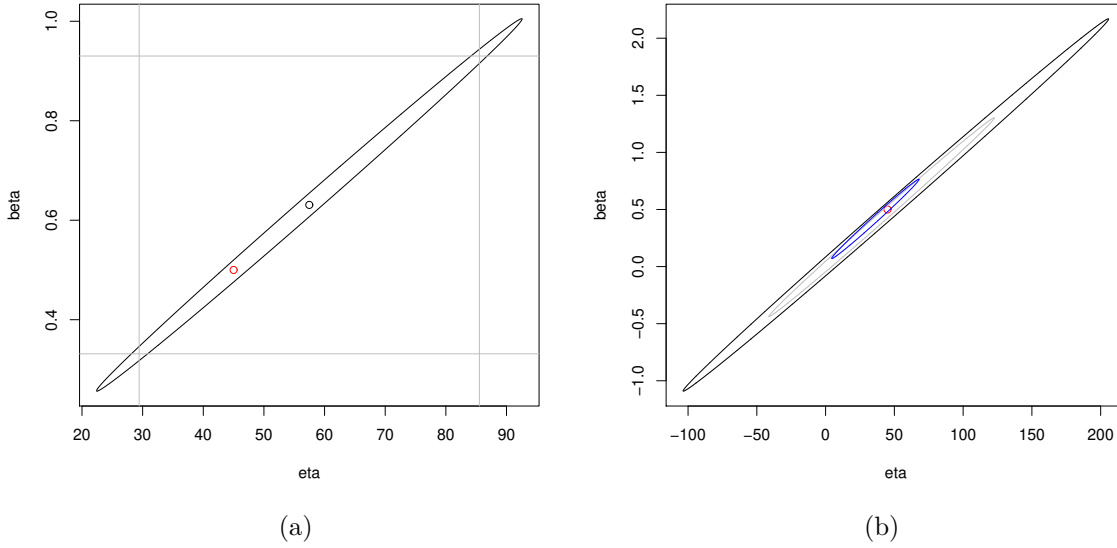


Figure 2: (a) shows the 95% joint confidence region for  $\beta$  and  $\eta$  from (2.32) and  $m = 10$  data sets generated using the parameter values in Example 2.6.2. The ellipse in the figure represents the 95% joint confidence region, while the grey lines represent the univariate 95% CIs for each of  $\beta$  and  $\eta$ . The black point marked in the figure is the point estimate for  $\beta$  and  $\eta$  and the red point represents the true value of  $\beta = 0.5$  and  $\eta = 45$ ; (b) shows the 95% joint confidence region for  $\beta$  and  $\eta$  using the parameter values in Example 2.6.3, with  $m = 1$  (black),  $m = 2$  (grey) and  $m = 5$  (blue). The red point represents the true value of  $\beta = 0.5$  and  $\eta = 45$ .

scheme, the Pseudo-likelihood method [3, 9], will be applied here to obtain estimators for  $\beta$ ,  $\eta$  and  $\sigma^2$ . We use the Euler method, which approximates the path of the process, so that the discretized form of the process has a likelihood that is usable and so can be maximised with respect to the parameter values.

### 3.1 Pseudo-MLE

The Euler scheme discretizes the process as (2.1). The increments  $I_{k+1} - I_k$  are conditionally independent Gaussian random variables with mean  $I_k(\beta N - \eta - \beta I_k)\Delta t$  and variance  $\sigma^2 I_k^2 (N - I_k)^2 \Delta t$ . Therefore the transition density of the process can be written as

$$p(I_{k+1}, (k+1)\Delta t | I_k, k\Delta t) = \frac{1}{\sqrt{2\pi\sigma^2 I_k^2 (N - I_k)^2 \Delta t}} \exp\left(-\frac{1}{2} \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t}\right), \quad (3.1)$$

where  $p(I_{k+1}, (k+1)\Delta t | I_k, k\Delta t)$  represents the conditional probability density that  $I[(k+1)\Delta t] = I_{k+1}$  given that  $I(k\Delta t) = I_k$ . Then a pseudo-likelihood is obtained as

$$L_n(\phi) = \prod_{k=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2 I_k^2 (N - I_k)^2 \Delta t}} \exp\left(-\frac{1}{2} \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t}\right) \right). \quad (3.2)$$

Taking the logarithm of (3.2) we have the log pseudo-likelihood

$$l_n(\phi) = -\frac{1}{2} \sum [\ln(2\pi\Delta t) + \ln\sigma^2 + \ln I_k^2 + \ln(N - I_k)^2] - \frac{1}{2} \sum \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t}. \quad (3.3)$$

The corresponding partial derivatives with respect to  $\beta$ ,  $\eta$  and  $\sigma^2$  are

$$\frac{\partial l_n(\phi)}{\partial \beta} = - \sum \frac{I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t} \cdot (-I_k N \Delta t + I_k^2 \Delta t), \quad (3.4)$$

$$\frac{\partial l_n(\phi)}{\partial (\eta)} = - \sum \frac{I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t} \cdot I_k \Delta t, \quad (3.5)$$

$$\frac{\partial l_n(\phi)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{(\sigma^2)^2} \cdot \frac{1}{2\Delta t} \cdot \sum \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{I_k^2 (N - I_k)^2}. \quad (3.6)$$

By setting all the partial derivatives equal to zero and solving these simultaneously, we find  $\hat{\beta}$ ,  $\hat{\eta}$  and  $\hat{\sigma}^2$  where the pseudo-likelihood function changes direction. We find that  $\hat{\beta}$ ,  $\hat{\eta}$  have the same expressions as the least squares estimators in (2.7) and (2.8), while  $\hat{\sigma}^2$  is almost the same as the least squares estimator (2.14) except that it has  $n$  in the denominator instead of  $(n-2)$ . We notice that  $\hat{\beta}$ ,  $\hat{\eta}$  and  $\hat{\sigma}^2$  are a unique solution to the partial derivative equations (3.4), (3.5) and (3.6), and the likelihood function (3.2) tends to 0 at the boundary. Since the values of the likelihood function are always positive, we

conclude that the turning point  $(\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$  maximises the pseudo-likelihood function (3.2). Therefore  $\hat{\phi} = (\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$  are the pseudo-MLEs for (1.2).

In the following sections we construct joint confidence regions for the pseudo-MLEs that we have obtained.

### 3.2 Exact Joint Confidence Region

We know that the pseudo-MLEs are exactly the same as the least squares estimators, except for a minor difference in the estimation of  $\sigma^2$ . If we want to find a joint  $100(1-\alpha)\%$  confidence region for  $\theta = (\beta, \eta)$  then we have already found this in the least squares case in (2.25) and (2.32) (for both  $m = 1$  and  $m > 1$ ) by obtaining an exact 95% confidence region for  $\theta$  as

$$(\theta - \hat{\theta})^T \left( \text{var}(\hat{\beta}, \hat{\eta}) \right)^{-1} \sigma^2 (\theta - \hat{\theta}) \leq \sigma^2 \chi_{\alpha,2}^2, \quad (3.7)$$

where  $\chi_{\alpha,2}^2$  is the upper  $\alpha$  point of the  $\chi^2$  distribution on 2 degrees of freedom, and then estimating  $\sigma^2$  by  $\hat{\sigma}^2$ . Note that we use  $\hat{\sigma}^2$  in (2.14) instead of  $\hat{\sigma}^2$  from the pseudo-MLE since the least squares estimator for  $\sigma^2$  is unbiased and is slightly better than the pseudo-MLE. Arnold (1998) [1] argues that if plug-in estimates are used for the variance, it is sensible to change the distribution from  $\chi_2^2$  to  $2F_{2,n-2}$  [4], to balance out the loss of accuracy because of the substitution that increases the area of the region. We replace  $\sigma^2$  by  $\hat{\sigma}^2$  and therefore it is more sensible to use  $2F_{2,n-2}$  here. Then it will lead to the same analytic form of the  $100(1-\alpha)\%$  joint confidence region for  $\beta$  and  $\eta$  as the least squares case in (2.24).

We already know the exact confidence region for  $\beta$  and  $\eta$  but we did not obtain the confidence region for all three pseudo-MLEs. In the following sections we construct large sample  $100(1-\alpha)\%$  joint confidence regions for all three pseudo-MLEs and for  $\beta$  and  $\eta$  as well for purposes of comparison. There are two ways to construct the asymptotic joint confidence region. The first method is based on the assumption that the pseudo-MLEs are approximately multivariate normally distributed, while the second is based on the likelihood ratio test statistic.

### 3.3 Asymptotic joint confidence regions based on the approximate multivariate normality of pseudo-MLEs

We can regard one data point as  $\mathbf{X} = (x_0, x_1, \dots, x_n)$ , which is a complete run with the initial data  $I_0$  and the transition probability as in (3.1). If we obtain  $m$  data points  $\mathbf{X} = (x_0, x_1, \dots, x_n)$  all with the same initial value and the same transition probability, then our  $m$  observations are independently and identically distributed and all with the pseudo-likelihood function as given in (3.2). Within this framework, we can apply the asymptotic maximum likelihood theory.

If  $m = 1$  (i.e. we have only one run) or  $m$  is very small this is not very helpful as the asymptotic theory requires the number of observations ( $m$  here) to be very large in order to be valid. In this case the estimators  $\hat{\beta}$  and  $\hat{\eta}$  are exactly the same as in the least squares case. So if  $\sigma^2$  is known their distribution is still exactly multivariate normal. Therefore if  $m = 1$  we use the exact confidence region as in (2.24) to estimate  $\beta$  and  $\eta$  for

the pseudo-MLE case. If  $m$  is large, we can then use the asymptotic pseudo-MLE theory and the likelihood ratio test which we will introduce in the next section.

First we find a joint confidence region for  $\boldsymbol{\phi} = (\beta, \eta, \sigma^2)$ . It is a standard result that the maximum likelihood estimators  $(\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$  for  $(\beta, \eta, \sigma^2)$  are approximately multivariately normally distributed with mean  $\boldsymbol{\phi}$  and variance  $\frac{1}{m}\Sigma(\boldsymbol{\phi})^{-1}$ , where  $\Sigma(\boldsymbol{\phi})$  is the Fisher information matrix defined in (3.9) [1, 16, 20], i.e.

$$\boldsymbol{\phi}_{(m)} \sim N^{(3)}\left(\boldsymbol{\phi}, \frac{1}{m}\Sigma(\boldsymbol{\phi})\right) \text{ approximately,} \quad (3.8)$$

where

$$\Sigma^{-1}(\boldsymbol{\phi}) = \sigma^{ij}(\boldsymbol{\phi}) = -E\left\{\frac{\partial^2}{\partial\phi_i\partial\phi_j}\ln f(\mathbf{X}; \boldsymbol{\phi})\right\}. \quad (3.9)$$

Here

$$f(\mathbf{X}; \boldsymbol{\phi}) = \prod_{k=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2 I_k^2 (N - I_k)^2 \Delta t}} \exp\left(-\frac{1}{2} \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t}\right) \right),$$

so that

$$\begin{aligned} \ln f(\mathbf{X}; \boldsymbol{\phi}) &= -\frac{1}{2} \sum [\ln(2\pi\Delta t) + \ln\sigma^2 + \ln I_k^2 + \ln(N - I_k)^2] \\ &\quad - \frac{1}{2} \sum \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t}. \end{aligned}$$

The regularity conditions required are that

$$\frac{d}{d\boldsymbol{\phi}} \int_{\Omega} f(\mathbf{x}; \boldsymbol{\phi}) d\mathbf{x} = \int_{\Omega} \frac{\partial}{\partial\boldsymbol{\phi}} f(\mathbf{x}; \boldsymbol{\phi}) d\mathbf{x} \quad (3.10)$$

where  $\Omega$  denotes the sample space. This property follows from the conditional normal distribution of the increments  $I_{k+1} - I_k$ . For example

$$\frac{\partial}{\partial\beta} \int_{\Omega} f(\mathbf{x}; \boldsymbol{\phi}) d\mathbf{x} = 0$$

as the integral over the sample space is one. But

$$\frac{\partial}{\partial\beta} f(\mathbf{x}; \boldsymbol{\phi}) = \sum_{k=1}^n \theta_k f(\mathbf{x}; \boldsymbol{\phi}), \text{ where } \theta_k = \frac{I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t}{\sigma^2 I_k^2 (N - I_k)^2 \Delta t} \cdot (I_k N \Delta t - I_k^2 \Delta t),$$

for  $k = 1, 2, 3, \dots, n$ .

$$\begin{aligned} \text{So } \int_{\Omega} \frac{\partial}{\partial\beta} f(\mathbf{x}; \boldsymbol{\phi}) d\mathbf{x} &= \sum_{k=1}^n E(\theta_k), \\ &= \sum_{k=1}^n E\{E(\theta_k | I_0, I_1, \dots, I_k)\}, \\ &= 0, \end{aligned}$$

as the conditional distribution of  $I_{k+1} - I_k$  given  $I_0, I_1, \dots, I_k$  is Gaussian with mean  $I_k(\beta N - \eta - \beta I_k)\Delta t$  and variance  $\sigma^2 I_k^2(N - I_k)^2 \Delta t$ . The other two parts of (3.10) corresponding to  $\eta$  and  $\sigma^2$  follow similarly.

The quadratic form associated with (3.8),

$$U = \sum_{i=1}^3 \sum_{j=1}^3 m \sigma^{ij}(\boldsymbol{\phi})(\hat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i)(\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j) \quad (3.11)$$

has an approximate chi-square distribution with three degrees of freedom for large  $m$ .

Because  $\hat{\boldsymbol{\phi}}$  is a strongly consistent estimate of  $\boldsymbol{\phi}$ , the statistics  $U$  will still have an asymptotic chi-square distribution with  $\sigma^{ij}(\boldsymbol{\phi})$  being substituted by  $\sigma^{ij}(\hat{\boldsymbol{\phi}})$ .

This will give a three dimensional confidence region for  $\boldsymbol{\phi} = (\beta, \eta, \sigma^2)$ . To actually evaluate this asymptotic confidence region for our case is very complicated. The equation (3.9) is very difficult to calculate since it involves the approximation of  $E\left(\frac{1}{(N-I_k)^2}\right)$ , and also it will bring in extra error from the approximation, so we do not use this confidence region in our examples.

On the other hand we could assume that  $\sigma$  is known and that we are trying to estimate  $\boldsymbol{\theta} = (\beta, \eta)$ . This is parallel to the estimation procedure that we used in the least squares problem (estimating  $\sigma$  by  $\hat{\sigma}$  and getting a two dimensional confidence region for  $\beta$  and  $\eta$ ). Then

$$\boldsymbol{\theta}_{(m)} \sim N^{(2)}\left(\boldsymbol{\theta}, \frac{1}{m}\Sigma(\boldsymbol{\theta})\right) \text{ approximately,}$$

where

$$\Sigma^{-1}(\boldsymbol{\theta}) = \sigma^{ij}(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} \ln f(\mathbf{X}; \boldsymbol{\theta})\right\}. \quad (3.12)$$

The associated quadratic form

$$U' = \sum_{i=1}^2 \sum_{j=1}^2 m \sigma^{ij}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \quad (3.13)$$

has an approximate chi-square distribution with two degrees of freedom for large  $m$ .

Note that  $\hat{\boldsymbol{\theta}}$  is the pseudo-MLE  $\hat{\boldsymbol{\theta}}(\sigma) = (\hat{\beta}(\sigma), \hat{\eta}(\sigma))$  with  $\sigma$  known and solves

$$\frac{\partial}{\partial\beta} \ln L_n(\boldsymbol{\theta}) = 0 \quad \text{and} \quad \frac{\partial}{\partial\eta} \ln L_n(\boldsymbol{\theta}) = 0.$$

Here  $L_n(\boldsymbol{\theta})$  is given by (3.2) except that  $\sigma$  is regarded as known.

If  $\sigma$  is actually unknown, we can substitute  $\sigma$  by its least squares estimator  $\hat{\sigma}$ . Then the distribution for that statistic  $U'$  is  $2F_{2, mn-2}$  [4]. We should use  $mn - 2$  here rather than  $m - 2$  as the estimator  $\hat{\sigma}^2$  is the average of  $mn - 2$  sums of squares. Also we should use the least squares estimator, not the pseudo-MLE for  $\hat{\sigma}$ , for the same reason as in section 3.2, although the results using the pseudo-MLE will be very close. If  $m$  is large,  $2F_{2, mn-2}$  will be approximately the same as a chi-square distribution with two degrees of freedom and the asymptotic confidence region will then approach the exact confidence region.

We are unable to work out this asymptotic confidence region numerically for the same reason as in the three dimensional case.

### 3.4 Joint confidence regions based on the likelihood ratio statistic

Another approximate confidence region is based on the likelihood ratio test statistic [1]. Suppose that we have  $m$  independent observations  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$  with common density  $f(\mathbf{X}|\phi)$ . Then we can approximate the  $100(1 - \alpha)\%$  confidence region for  $\phi$  by

$$(\phi : -2\log R_n(\phi) < \chi_{3,1-\alpha}^2). \quad (3.14)$$

Here

$$R_n(\phi) = \frac{L_m(\phi)}{L_n(\hat{\phi})},$$

where the vector  $\hat{\phi}$  contains the pseudo-MLEs for  $\phi$ , the parameters,

$$L_m(\phi) = \prod_{j=1}^m L_{n,j}(\phi) \quad (3.15)$$

and  $L_{n,j}(\phi) =$

$$\prod_{k=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t}} \exp \left( -\frac{1}{2} \frac{[I_{k+1,j} - I_{k,j} - I_{k,j}(\beta N - \eta - \beta I_{k,j})\Delta t]^2}{\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t} \right) \right).$$

So a  $100(1 - \alpha)\%$  confidence region for  $\phi$  is

$$\sum_{j=1}^m \sum_{k=1}^n \left( \ln \frac{2}{\sqrt{2\pi\hat{\sigma}^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t}} - \ln \frac{2}{\sqrt{2\pi\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t}} + \frac{[I_{k+1,j} - I_{k,j} - I_{k,j}(\beta N - \eta - \beta I_{k,j})\Delta t]^2}{\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t} \right) - mn < \chi_{\alpha,3}^2.$$

Again if  $\sigma^2$  is assumed known, a similar argument shows that a  $100(1 - \alpha)\%$  confidence region for  $\theta$  is

$$\sum_{j=1}^m \sum_{k=1}^n \left( \frac{[I_{k+1,j} - I_{k,j} - I_{k,j}(\beta N - \eta - \beta I_{k,j})\Delta t]^2}{\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t} - \frac{[I_{k+1,j} - I_{k,j} - I_{k,j}(\hat{\beta} N - \hat{\eta} - \hat{\beta} I_{k,j})\Delta t]^2}{\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t} \right) < \chi_{\alpha,2}^2. \quad (3.16)$$

Here again  $\hat{\theta}$  is the pseudo-MLE  $\hat{\theta}(\sigma) = (\hat{\beta}(\sigma), \hat{\eta}(\sigma))$  with  $\sigma$  known, and solves

$$\frac{\partial}{\partial \beta} \ln L_m(\theta) = 0 \quad \text{and} \quad \frac{\partial}{\partial \eta} \ln L_m(\theta) = 0.$$

In these equations  $L_m(\theta)$  is given by (3.15) but regarded as a function of  $\theta = (\beta, \eta)$  with  $\sigma$  known rather than as a function of  $\phi = (\beta, \eta, \sigma)$ .

Again if we replace the unknown  $\sigma$  by  $\hat{\sigma}$  (the least squares estimator), then the distribution should be  $2F_{\alpha,2,mn-2}$ .

Then (3.16) can be written as

$$\sum_{j=1}^m \sum_{k=1}^n \left( \frac{[I_{k+1,j} - I_{k,j} - I_{k,j}(\beta N - \eta - \beta I_{k,j})\Delta t]^2}{\sigma^2 I_{k,j}^2 (N - I_{k,j})^2 \Delta t} - (mn - 2) \right) < 2F_{\alpha,2,mn-2},$$

which is equivalent to

$$\begin{aligned} \frac{1}{\hat{\sigma}^2} \left( \sum_{j=1}^m \sum_{k=1}^n \frac{(I_{k+1,j} - I_{k,j})^2}{I_{k,j} (N - I_{k,j})^2 \Delta t} + \beta^2 mn \Delta t + \eta^2 \sum_{j=1}^m \sum_{k=1}^n \frac{\Delta t}{(N - I_{k,j})^2} + \eta \beta \sum_{j=1}^m \sum_{k=1}^n \frac{-2\Delta t}{N - I_{k,j}} \right. \\ \left. + \beta \sum_{j=1}^m \sum_{k=1}^n \frac{-2(I_{k+1,j} - I_{k,j})}{I_{k,j}(N - I_{k,j})} + \eta \sum_{j=1}^m \sum_{k=1}^n \frac{2(I_{k+1,j} - I_{k,j})}{I_{k,j}(N - I_{k,j})^2} \right) - (mn - 2) < 2F_{\alpha,2,mn-2}. \end{aligned}$$

This can be simplified as

$$\begin{aligned} \sum_{j=1}^m \sum_{k=1}^n y_{k,j}^2 + \beta^2 nm \Delta t + \eta^2 \sum_{j=1}^m \sum_{k=1}^n u_{k,j}^2 + \eta \beta 2\sqrt{\Delta t} \sum_{j=1}^m \sum_{k=1}^n u_{k,j} + \beta(-2\sqrt{\Delta t}) \sum_{j=1}^m \sum_{k=1}^n y_{k,j} \\ + \eta(-2) \sum_{j=1}^m \sum_{k=1}^n u_{k,j} y_{k,j} - (mn - 2)\hat{\sigma}^2 < 2\hat{\sigma}^2 F_{\alpha,2,mn-2}, \end{aligned}$$

with  $u_k$  and  $y_k$  defined in (2.2).

This can be written as

$$mn \Delta t \left( \beta - \hat{\beta} \right)^2 + \sum_{j=1}^m \sum_{k=1}^n u_{j,k}^2 (\eta - \hat{\eta})^2 + 2\sqrt{\Delta t} \sum_{j=1}^m \sum_{k=1}^n u_{k,j} (\beta - \hat{\beta})(\eta - \hat{\eta}) < 2\hat{\sigma}^2 F_{\alpha,2,mn-2} - D, \quad (3.17)$$

where

$$D = -mn \Delta t \hat{\beta}^2 - \sum_{j=1}^m \sum_{k=1}^n u_{j,k}^2 \hat{\eta}^2 - 2\sqrt{\Delta t} \sum_{j=1}^m \sum_{k=1}^n u_{k,j} \hat{\beta} \hat{\eta} + \sum_{j=1}^m \sum_{k=1}^n y_{k,j}^2 - (mn - 2)\hat{\sigma}^2.$$

The region (3.17) has the same form as the exact confidence region (2.32) apart from the subtraction of a constant  $D$  on the right hand side. We have shown that (2.33) is strictly negative and therefore the  $100(1 - \alpha)\%$  confidence region for  $\theta$  (3.17) is an ellipse centered at the pseudo-MLE  $\hat{\beta}$  and  $\hat{\eta}$ . We numerically compare the exact 95% confidence region for  $\theta$  with the asymptotic confidence region obtained by using the likelihood ratio test in the following example, and establish the size of the difference  $D$  in this case.

**Example 3.4.1** Assume that the parameters are given by  $T = 5$ ,  $I(0) = 10$ ,  $\beta = 0.5$ ,  $\mu = 20$ ,  $\gamma = 25$ ,  $N = 100$ ,  $m = 100$  and  $\sigma^2 = 0.03$  for the model (1.2).

We use the same method as in Example 2.6.2 to simulate  $m = 100$  data sets and sample from them. With the samples we obtained, we calculate both the 95% joint confidence region for  $\beta$  and  $\eta$  (2.32) and the asymptotic confidence region obtained using the likelihood ratio test (3.17). These confidence regions are shown in Figure 3. The two confidence regions are almost identical and the difference between them can hardly be seen. We calculated  $D$ , the difference between the two confidence regions as shown in (3.17), which is very small,  $-2.373 \times 10^{-12}$  in this case.

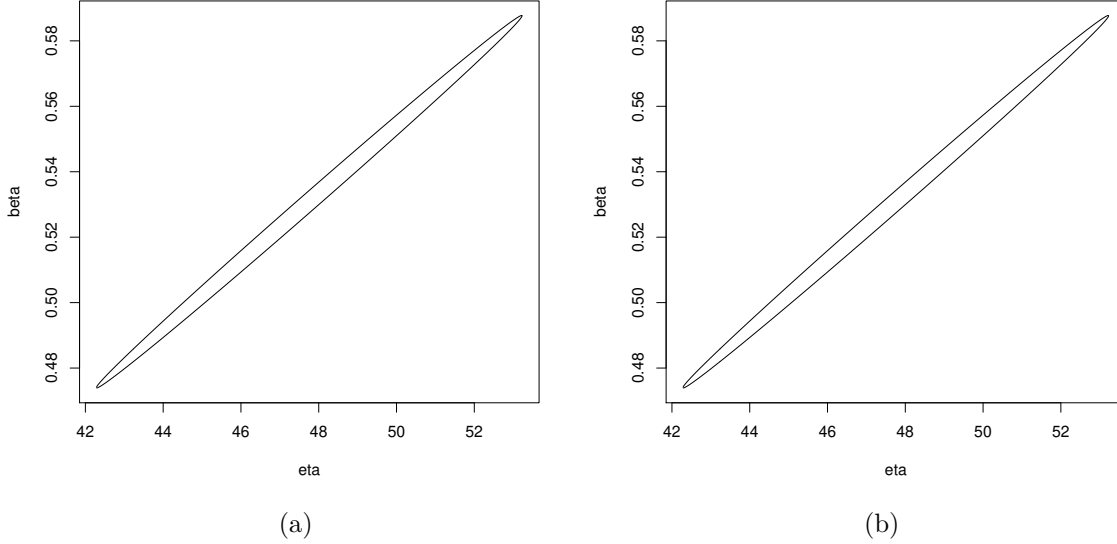


Figure 3: (a) shows the exact 95% joint confidence region for  $\beta$  and  $\eta$  (2.32) using the parameter values in Example 3.4.1; (b) shows the approximate likelihood ratio based confidence region using (3.17).

## 4 Summary and Further Work

In this paper we have applied the pseudo-MLE and the least squares method to estimate the parameters in the stochastic SIS model. For the least squares method, we started with the case in which only one data set is available and then improved our method by considering the case where more than one data set is available. We have obtained the point estimators,  $100(1 - \alpha)\%$  CIs and  $100(1 - \alpha)\%$  joint confidence regions for  $\beta$  and  $\eta$  for both cases. We also investigated which factors influence the width of the CIs and the areas of the confidence regions. Theorem 2.2 states that the asymptotic widths of the CIs for both  $\beta$  and  $\eta$  strictly decrease as the total time period  $T$  increases and do not depend on the size of the time step  $\Delta t$ . Example 2.6.1 shows that a sample from one location with denser observations does not give narrower CIs, while using more than one sample taken at different locations and getting a sample with a longer period of observation at one location decreases the width of CIs significantly and improves the efficiency of estimation. Examples 2.5.2 and 2.6.3 show that the area of the confidence region decreases with increasing total time period  $T$  and increasing number of samples  $m$ .

We have also obtained pseudo-MLEs which are almost the same as the point estimators from the least squares case, with a minor difference in the estimators of  $\sigma^2$ . For obtaining the confidence region for the pseudo-MLEs we considered the following two cases. When the number of samples  $m$  is small, we obtained the exact confidence region for  $\beta$  and  $\eta$  in the pseudo-MLE case based on the least squares method. When  $m$  is large, we used the asymptotic MLE theory and the likelihood ratio test to obtain the large sample confidence regions for both the three dimensional case (using all three pseudo-MLEs) and the two dimensional case (estimating  $\beta$  and  $\eta$  assuming that  $\sigma$  is known). We only



calculated numerically the asymptotic confidence region based on the likelihood ratio test for  $\beta$  and  $\eta$ . Example 3.4.1 shows that the numerical asymptotic confidence region using the likelihood ratio test for  $\beta$  and  $\eta$  is almost identical to the exact confidence region.

Comparing the least squares estimation method and the pseudo-MLE method, we find that although the pseudo-MLE is more popular for parameter estimation for SDEs, least squares estimation gave the same point estimators and joint confidence region as the pseudo-MLE and is easier to apply. In our case least squares estimation is advantageous.

The Bayesian approach is another popular way to estimate the parameters for SDEs [9]. In practice, we often have some information about the value of parameters before data is collected. The Bayesian approach is advantageous in this circumstance since it includes the prior information about the model parameters in the form of one or more prior distributions [5]. We are applying the Bayesian approach to our stochastic SIS model and will report the results in a further paper.

## References

- [1] Arnold, B.C. and Shavelle, R.M., Joint confidence sets for the mean and variance of a Normal distribution, *The American Statistician*, 52(2)(1998), 133-140.
- [2] Bishwal, J.P.N., *Parameter Estimation in Stochastic Differential Equations*, Berlin/Heidelberg: Springer-Verlag, 2008.
- [3] Cao, J. and Hu, L., Asymptotic properties of a pseudo-MLE for CIR model, Proceedings of the 2010 International Conference on IIGSS-CPS, Nanjing, China, Vol. 1: Advances on Probability and Statistics, 28-31 (2010), 206-209.
- [4] Douglas, J.B., Confidence regions for parameter pairs, *The American Statistician*, 41(1) (1993), 43-45.
- [5] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., *Bayesian Data Analysis*, London: Chapman and Hall, 1995.
- [6] Gray, A., Greenhalgh, D., Hu, L., Mao, X. and Pan, J., A stochastic differential equation SIS epidemic model, *SIAM Journal of Applied Mathematics*, 71(3) (2011), 876-902.
- [7] Gushchin, A.A. and K uchler, U., Asymptotic properties of maximum-likelihood-estimators for a class of linear stochastic differential equations with time delay, *Bernoulli*, 5 (2000), 1059-1098.
- [8] Hethcote, H.W. and Yorke, J.A., *Gonorrhoea transmission dynamics and control*, Lecture Notes in Biomathematics 56, Springer-Verlag, 1994.
- [9] Iacus, S.M., *Simulation and Inference for Stochastic Differential Equations with R Examples*, New York: Springer, 2008.
- [10] Stuart, A. and Ord, J.K., *Kendall's Advanced Theory Of Statistics: Volume 2, Classical Inference and Relationship*, 5th Edition, London: Edward Arnold, 1991.

- [11] Kristensen, N.R., Madsen, H. and Young, P.C., Parameter estimation in stochastic grey-box model, *Automatica*, 40 (2004), 225-237.
- [12] KÜchler, U. and Kutoyants, Y., Delay estimation for some stationary diffusion-type processes, *Scandinavian Journal of Statistics*, 27 (2000), 405-414.
- [13] KÜchler, U. and Sorensen, M., A simple estimator for discrete-time samples from affine stochastic delay differential equations, *Statistical Inference for Stochastic Processes*, 13(2) (2010), 125-132.
- [14] KÜchler, U. and Vasil'jev, V.A., Sequential identification of linear dynamic systems with memory, *Statistical Inference for Stochastic Processes*, 8 (2005), 1-24.
- [15] Lamb, K.E., Greenhalgh, D. and Robertson, C., A simple mathematical model for genetic effects in pneumococcal carriage and transmission, *Journal of Computational and Applied Mathematics*, 235(7) (2011), 1812-1818.
- [16] Lindsey, J.K., *Parametric Statistical Inference*, Oxford: Clarendon Press, 1996.
- [17] Lipsitch, M., Vaccination against colonizing bacteria with multiple serotypes, *Proceedings of the National Academy of Sciences*, 94 (1997), 6571-6576.
- [18] Mao, X., *Stochastic Differential Equations and Applications*, Chichester: Horwood Publishing, 2nd Edition, 2007.
- [19] Mao, X. and Yuan, C., *Stochastic Differential Equations with Markovian Switching*, London: Imperial College Press, 2006.
- [20] Mood, A.M., *Introduction to the Theory of Statistics*, New York: McGraw-Hill, 1950.
- [21] Nielsen, J.N., Madsen, H. and Young, P.C., Parameter estimation in stochastic differential equations: an overview, *Annual Reviews in Control*, 24 (2000), 83-94.
- [22] Rawlings, J.O., *Applied Regression Analysis: a Research Tool*, Belmont, CA: Wadsworth, 1988.
- [23] Reiß, M., Minimax rate for nonparametric drift estimation in affine stochastic delay differential equations, *Statistical Inference for Stochastic Processes*, 5 (2002), 131-152.
- [24] Reiß, M., *Nonparametric Estimation for Stochastic Delay Differential Equations*, PhD thesis, Institut für Mathematik, Humboldt-Universität zu Berlin, 2002.
- [25] Reiß, M., Adaptive estimation for affine stochastic delay differential equations, *Bernoulli*, 11 (2005), 67-102
- [26] Timmer, J., Parameter estimation in nonlinear stochastic differential equations, *Chaos, Solitons and Fractals*, 11, 2571-2578.
- [27] Young, P., Parameter estimation for continuous-time models - a survey, *Automatica*, 17(1) (1981), 23-39.