Maximum Likelihood Estimation for Itô Stochastic SIR Models

R. Faith and G.M. Kaufman

April 5, 2023

Easy to compute exact maximum likelihood estimators (MLEs) for parameters of a stochastic bivariate Itô Susceptible-Infected-Recovered (SIR) model and for parameters of an extension that treats undercounting are presented here.

KEY WORDS: Infectious disease modeling, stochastic SIR model, Maximum likelihood

1. Introduction

Properties of exact maximum likelihood estimators of parameters of a discrete version of an Itô stochastic SIR model are documented. Euler's method leads to a discrete time bivariate stochastic auto-regressive data generating process whose parameters are estimable using standard MLE methods. Monte Carlo simulations indicate that this MLE is well behaved.

Emerging experience with SARS Covid data suggests that the standard Itô SIR model be extended to incorporate undercounting, Introduction of undercounting, accompanied with methods for parameter estimation appear in Sect. 5. Observed counts of infected individuals are modeled as a fixed proportion of true infected individual counts. This Occam's Razor assumption simplifies computation but captures the essence of undercounting. Tables 4.1 and 5.1 suggest that application to an early sample in the life cycle of an epidemic produces sensible projections of its future path.

The literature describing maximum likelihood estimation for non-stationary univariate Itô processes is extensive (Lo (2008), Aït Sahalia (2002) for example) but less so for multivariate non-stationary Itô processes. The rather simple functional form of the SIR model permits calculation of easy to compute MLEs for parameters of the bivariate stochastic version studied here.

2. SIR Model Assumptions and Properties

The SIR epidemic model partitions a population of N individuals into three mutually exclusive sets: a set of infected individuals, a set of susceptible individuals and a set of immune individuals possessing counts $I(t)$, $S(t)$ and $R(t)$ at time t respectively. Equations governing disease dynamics are

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta}{N} S(t)I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t) \end{aligned} \tag{2.1}$$

subject to

$$S(t) + I(t) + R(t) = N. \tag{2.2}$$

The SIR Model for Spread of Disease

Taken together (2.1) and (2.2) imply

$$\frac{dI(t)}{dt}=\frac{\beta}{N}S(t)I(t)-\gamma I(t). \tag{2.3}$$

Initial conditions $S(0)=S_0, I(0)=I_0$ and $R(0)=N-S_0-I_0$. The sign of $\frac{\beta S_0}{N\gamma}-1$ determines the behavior of Eqns. (2.1), (2.2) and (2.3). If $\frac{\beta S_0}{N\gamma}<1$ disease does not propagate and if $\frac{\beta S_0}{N\gamma}>1$ disease propagates. The intuition behind defining the *mean infection rate* to be $R_0=\frac{\beta S_0}{N\gamma}$ is that it is a product of the disease transmission rate β per unit time, a mean infection time $1/\gamma$ and an initial condition $\frac{S_0}{N}\in(0,1)$. When $S_0/N\approx 1$, R_0 is close to the product of the transmission rate and mean infection time. See Allen (2017)) van den Driessche (2017).

3. Itô SIR and Euler's Method

The Itô SIR model corresponding to Eqns. (2.1) and (2.2) is a pair of stochastic differential equations shown below. Application of Euler's Method to Eqns. (2.1) and (2.2) yields a bivariate discrete non-linear auto regressive time series in susceptible and infected counts. Allen (2017)

3.1 Stochastic SIR Model

Itô SDE equations for the SIR Model are:

$$dS(t)=-\frac{\beta}{N}I(t)S(t)dt-\sqrt{\frac{\beta}{N}I(t)S(t)}\times dW_1(t) \tag{3.1a}$$

and

$$dI(t)=[\frac{\beta}{N}I(t)S(t)-\gamma I(t)]dt+\sqrt{\frac{\beta}{N}I(t)S(t)}\times dW_1(t)-\sqrt{\gamma I(t)}\times dW_2(t) \tag{3.1b}$$

Here $W_1(t)$ and $W_1(t)$ are (mutually) independent Wiener processes distributed as $N(0,dt)$ (Normal mean zero, variance dt). Although domains of $W_1(t)$ and that $W_2(t)$ are $(-\infty,\infty)$ the process stops at time $\tau=\inf t\in(0,\infty)$ such that $I(t)\le 0$ almost surely. This stopping time feature complicates, but does not impede calculation of properties of parameter estimates.

3.2 Euler-Maruyama

Denominate time in small constant intervals Δt , $t = 0,1,...,T$ and rescale Δt to one. Euler-Maruyama formulae for Eqns. (2.1a), (2.1b)) and (2.3c) are

$$\begin{aligned} S_t &= S_{t-1} - \frac{\beta}{N} S_{t-1} I_{t-1} \\ I_t &= \frac{\beta}{N} S_{t-1} I_{t-1} + (1-\gamma) I_{t-1} \ , \qquad t = 0,1,...,T \\ R_t - R_{t-1} &= \gamma I_{t-1} \end{aligned} \tag{3.2}$$

To arrive at a discretized version of Eqns. (3.1), As in Allen (2017), define Gaussian independent increments $W_1(t+\Delta t) - W_1(t) \equiv \varepsilon_{1t} \Delta t$ and $W_2(t+\Delta t) - W_2(t) \equiv \varepsilon_{2t} \Delta t$. Append them to the first two equations in Eqns. (3.2).

$$\Delta S_t = S_t - S_{t-1} = -\frac{\beta}{N} S_{t-1} I_{t-1} - \sqrt{\frac{\beta}{N} S_{t-1} I_{t-1}} \varepsilon_{1t} \tag{3.3a}$$

$$\Delta I_t = I_t - I_{t-1} = \frac{\beta}{N} S_{t-1} I_{t-1} - \gamma I_{t-1} + \sqrt{\frac{\beta}{N} S_{t-1} I_{t-1}} \varepsilon_{1t} - \sqrt{\gamma I_{t-1}} \varepsilon_{2t} \ . \tag{3.3b}$$

Conditional expectations of ΔS_t and ΔI_t given S_{t-1} and I_{t-1} are $E(\Delta S_t \,|S_{t-1}, I_{t-1}) = -\frac{\beta}{N} S_{t-1} I_{t-1}$ and $E(\Delta I_t \,|S_{t-1}, I_{t-1}) = \frac{\beta}{N} S_{t-1} I_{t-1} - \gamma I_{t-1}$. The variance matrix of ΔS_t and ΔI_t given S_{t-1} and I_{t-1} is

.
$$Var(\begin{pmatrix} \Delta S_t \\ \Delta I_t \end{pmatrix} |S_{t-1}, I_{t-1}) = Var(\mathbf{G}_{t-1} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}) \ . \tag{3.4a}$$

Inspection of random terms in Eqns. (3.3a) and (3.3b) lead to a choice (Allen (2017)) of

$$\mathbf{G}_{t-1} = \begin{bmatrix} -\sqrt{\frac{\beta}{N} S_{t-1} I_{t-1}} & 0 \\ \sqrt{\frac{\beta}{N} S_{t-1} I_{t-1}} & -\sqrt{\gamma I_{t-1}} \end{bmatrix} \tag{3.4b}$$

so
$$Var(\begin{pmatrix} \Delta S_t \\ \Delta I_t \end{pmatrix} |S_{t-1}, I_{t-1}) = \begin{bmatrix} \frac{\beta}{N} S_{t-1} I_{t-1} & -\frac{\beta}{N} S_{t-1} I_{t-1} \\ -\frac{\beta}{N} S_{t-1} I_{t-1} & \frac{\beta}{N} S_{t-1} I_{t-1} + \gamma I_{t-1} \end{bmatrix} . \tag{3.4c}$$

3.1 Random Stopping

Eqns. (3.1a) and (3.1b) are subject to random stopping at $S(\tau)\le 0 \text{ or } I(\tau)\le 0 \text{ or both}$. That is, at the earliest time at which there are no infected individuals or no individual remains susceptible or both. This feature complicates computation of asymptotic properties of estimators but does not impede computation of a finite maximum likelihood estimator for a sample terminating at time $T<\tau$. (Sørensen (2008)).

Remark: In what follows description shuttles back and forth between continuous and discrete version notation.

4. Likelihood Functions and MLE

If the probability law governing the sequence of random variables $\mathbf{X}_t \equiv (\Delta S_t, \Delta I_{t-1}),\ t=1,2,....,T$ is the discretized version Eqns. (3.5) of Eqns. (3.2) then the joint probability of observing $\mathbf{X}_t \in \mathbf{dx}_t$ is a product of bivariate Normal densities. Define

$$\boldsymbol{\mu}_{t-1}=\begin{pmatrix}-\dfrac{\beta}{N}S_{t-1}I_{t-1}\\ \dfrac{\beta}{N}I_{t-1}-\gamma\end{pmatrix} \text{ and } \boldsymbol{\Sigma}_{t-1}=\frac{\beta}{N}S_{t-1}I_{t-1}\begin{bmatrix}1 & -1\\ -1 & 1+\dfrac{N\gamma}{\beta S_{t-1}}\end{bmatrix}\quad t=1,2,....,T. \tag{4.1}$$

Then $Prob\{\mathbf{X}_t\in\mathbf{dx}_t, t=1,..,T\}=\prod_{t=1}^{T} f_N(\mathbf{x}_t\,|\,\boldsymbol{\mu}_{t-1},\boldsymbol{\Sigma}_{t-1})\,\mathbf{dx}_t$ and the log likelihood function is

$$\ln \mathcal{L}(\frac{\beta}{N},\gamma\,|\,\mathbf{x}_1,...,\mathbf{x}_T)\propto\sum_{t=1}^{T}\ln|\boldsymbol{\Sigma}_{t-1}|-(\mathbf{x}_t-\boldsymbol{\mu}_{t-1})^t\boldsymbol{\Sigma}_{t-1}^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_{t-1}). \tag{4.2}$$

Define statistics

$$Z_{1T}^2=\frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta S_t+\Delta I_t)^2}{I_{t-1}},\quad Z_{2T}^2=\frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta S_t)^2}{S_{t-1}I_{t-1}},\quad \bar{I}_T=\frac{1}{T}\sum_{t=1}^{T}I_{t-1} \text{ and } \bar{U}_T=\frac{1}{T}\sum_{t=1}^{T}I_{t-1}S_{t-1}. \tag{4.3}$$

Assertion 4.1: For sample sizes less than an observed stopping time τ, a maximizer $\hat{\boldsymbol{\theta}}\equiv(\hat{\gamma},\hat{\beta})^t$ of $\ln \mathcal{L}(\frac{\beta}{N},\gamma\,|\,\mathbf{x}_1,...,\mathbf{x}_T)$ obtains at

$$\hat{\gamma}=\frac{-1+\sqrt{1+4\bar{I}_T Z_{1T}^2}}{2\bar{I}_T} \quad\text{and}\quad \hat{\beta}=N\times\frac{-1+\sqrt{1+4\bar{U}_T Z_{2T}^2}}{2\bar{U}_T}. \tag{4.4}$$

The matrix of second derivatives of $\ln \mathcal{L}$ is

$$\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln \mathcal{L} = \begin{bmatrix} \frac{1}{\gamma^3}(\gamma - 2Z_{1T}^2) & 0 \\ 0 & \frac{1}{\beta^3}(\beta - 2Z_{2T}^2) \end{bmatrix} \tag{4.5}$$

so the likelihood function is concave for values of $\gamma < 2Z_{1T}^2$ and $\beta < 2Z_{1T}^2$. In particular at $\hat{\boldsymbol{\theta}} \equiv (\hat{\gamma}, \hat{\beta})^t$

$$\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln \mathcal{L} = \begin{bmatrix} \frac{1}{\hat{\gamma}^3}(\hat{\gamma} - 2Z_{1T}^2) & 0 \\ 0 & \frac{1}{\hat{\beta}^3}(\hat{\beta} - 2NZ_{2T}^2) \end{bmatrix} < \mathbf{0}. \tag{4.6}$$

4.2 Monte Carlo Simulation of MLE Estimators

A Monte Carlo study uncovers features of MLEs $\hat{\gamma}$ and $\hat{\beta}$ of γ and β and sample statistics. Figures 4.1, 4.2 and 4.3 display simulated sample paths for population size 1000 and initial infection count 30. Empirical histograms of $\hat{\gamma}$ and $\hat{\beta}$ appear in Figures 4.4 and 4.5.

Figure 4.1

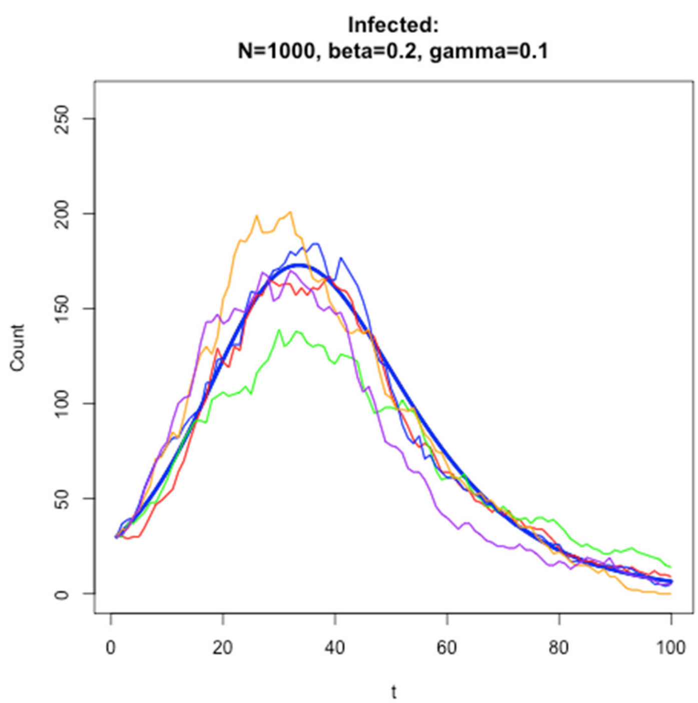


Figure 4.2

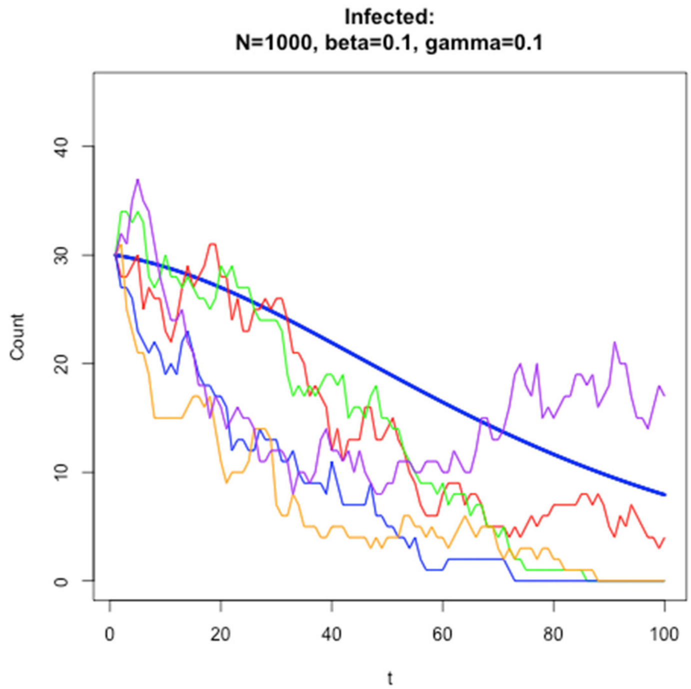
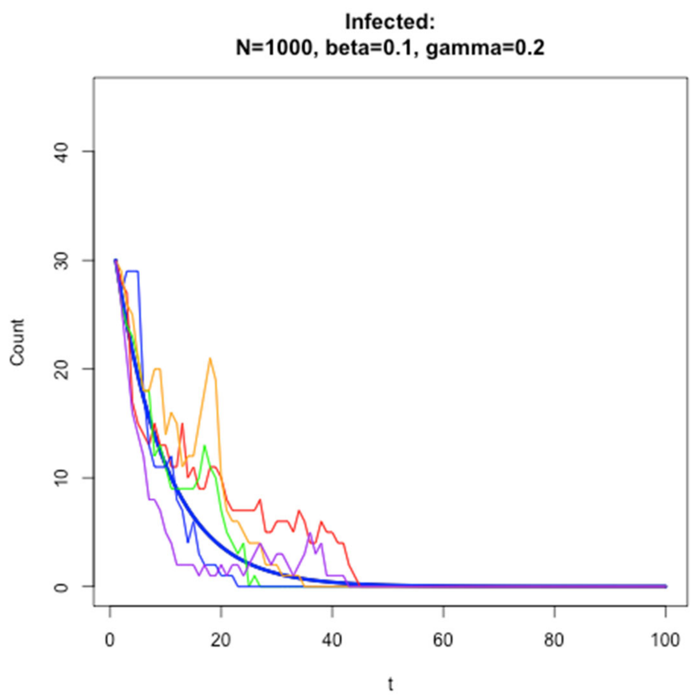


Figure 4.3



For $\gamma < \beta$ mean values of observables increase, peak and decline in accord with typical graphical presentations of SIR simulation output. For $\beta \le \gamma$ mean values of observables decline monotonically to extinction at zero.

2.1 Monte Carlo Statistics

Initial conditions are population size $N = 1000$, susceptible count S_0 at time zero and infection count $I_0 = 30$ at time zero along with specification of a simulation time interval $(0, T)$ .Here *sample size* is (the nearest integer to) a fraction $f \in (0,1)$ of T if extinction obtains at a time $\tau \ge T$ and is τ otherwise. MLEs of β and γ for the example displayed in Table 4.1 appear to be unbiased and reasonably "tight": Standard deviations are 0.0079 and 0.0042 respectively. Q-Q plots of β and γ show curvature away from a normal distribution in the tails, but not enough to compromise utility of standard confidence interval calculations in this particular case.

Table 4.1

$N = 1000,\ I(0) = 30 \quad \beta = 0.2 \quad \gamma = 0.1$

$Simulation\ Runs = 500 \quad Time\ Periods = 50$

	Beta	Gamma	Max Infected	Peak Time
Min	0.1660	0.0905	107	22
First Q	0.1946	0.0974	164	30
Median	**0.2000**	**0.1002**	**181**	**33**
Mean	**0.1998**	**0.1004**	**180**	**33**
Third Q	0.2052	0.1031	197	37
Max	0.2268	0.1146	245	50
Std Dev	0.0079	0.0042	23.409	5.203

Figure 4.4 Gamma MLEs

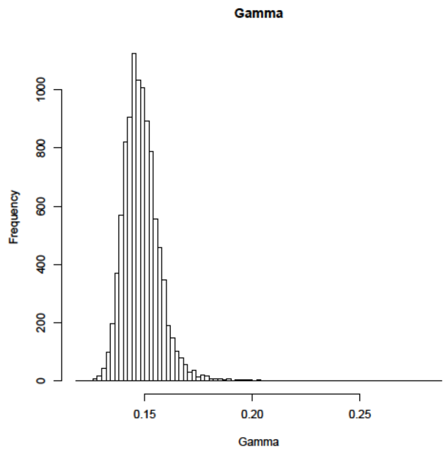


Figure 4.5 Beta MLEs

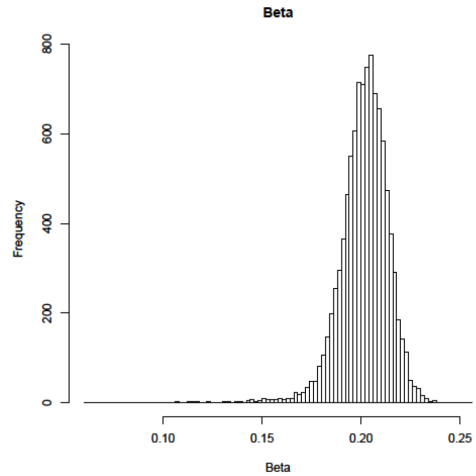


Figure 4.6 Beta vs Gamma Scatterplots

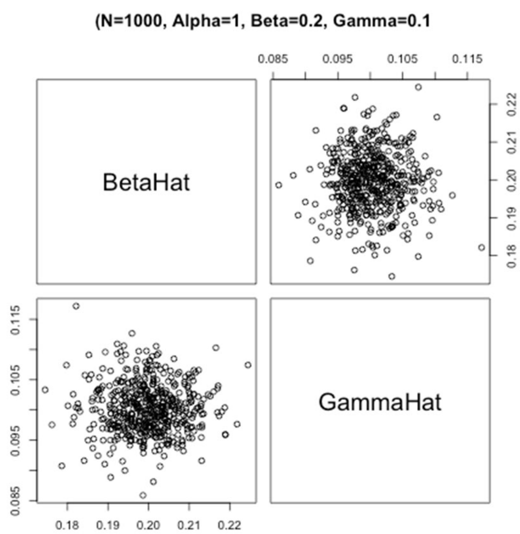


Figure 4.7 Beta Q-Q Plot

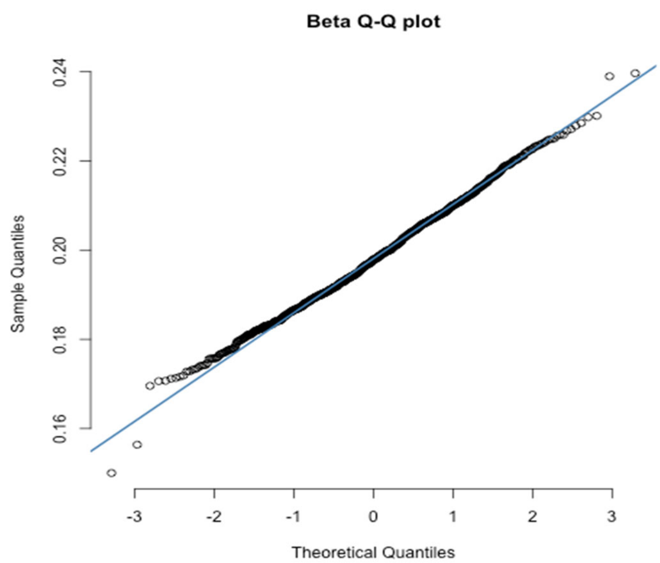
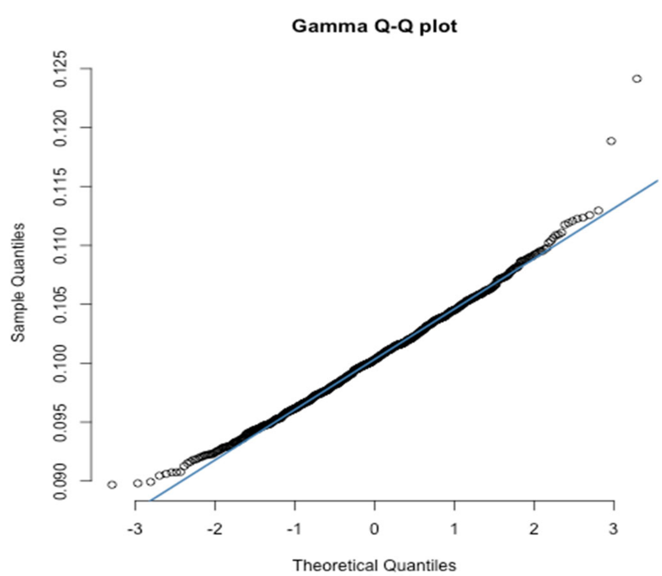


Figure 4.8 Gamma Q-Q Plot



5. Undercounts

Almost all current Covid19 epidemic data series are incomplete and do not accurately count infected individuals. While both over- and undercounting can occur, undercounting is the more likely scenario. To account for undercounting modify Eqns. (3.1a) and (3.1b). Partition the set of infected individuals at time t into two sets: identified infected individuals and infected individuals not identified. Define $M(t)$ to be the number of *identified* infected individuals at time t. The undercount is then $I(t)-M(t)$ and a model flow diagram is

$$S(t) \begin{matrix} \nearrow & M(t) & \searrow \\ \searrow & I(t)-M(t) & \nearrow \end{matrix} R(t) \quad . \tag{5.1}$$

Define $Y(t)=S(t)+I(t)$. Now $Y(t)-M(t)$ and $M(t)$ are observed in place of $S(t)$ and $I(t)$ .The observational flow diagram is

$$Y(T) \to M(t) \to R(t)\ . \tag{5.2}$$

When undercounting takes place *observed data* is representable as a matrix of (2×1) vectors $((S_1+I_1-M_1,M_1)^t,...,(S_1+I_1-M_1,M_1)^t)^t$ and *unobserved (latent) data*, a matrix of (2×1) vectors $((S_1,I_1)^t,...,(S_T,I_T)^t)^t$. The transformation from $(S_T,I_T)^t$ to $(Y_t,I_t)^t$ with $Y_t=S_t+I_t$ is one to one, so the vector $\mathbf{Z}=(\mathbf{Z}_1,...,\mathbf{Z}_T)^t$, $\mathbf{Z}_t \equiv (Y_t,I_t)$, is a sufficient (but not minimal) statistic. Statistics M_t and $Y_t=S_t+I_t$ are observational equivalents of $(S_t,I_t)^t$. Modeling each undercount as a fixed fraction of true infected individual counts is the simplest way to treat undercounting. It completes specification of the data generating process described in Eqns. (5.1) and (5.2).

Assumption: $M(t)$ is a fixed but unknown fraction α of $I(t)$. $M(t)=\alpha I(t)$, $t\in(0,T)$.

Replace $I(t)$ with $M(t)/\alpha$ in Eqn. (3.1b) and multiply both sides by α to arrive at

$$dM(t)=[\frac{\beta}{N}M(t)S(t)-\gamma M(t)]dt+\sqrt{\alpha\frac{\beta}{N}M(t)S(t)}\times dW_1(t)-\sqrt{\alpha\gamma M(t)}\times dW_2(t) \tag{5.3}$$

subject to $S(t)=Y(t)-\frac{M(t)}{\alpha}>0$. Define $S_t(\alpha)=Y_t-\frac{M_t}{\alpha}$. The discrete local mean of $(\Delta Y_t,\Delta M_t)^t$ is

$$\boldsymbol{\mu}_{t-1} = \begin{pmatrix} -\frac{\gamma M_{t-1}}{\alpha} \\ \frac{\beta S_{t-1}(\alpha)M_{t-1}}{N}-\gamma M_{t-1} \end{pmatrix} \tag{5.4}$$

and its discrete local variance is

$$\mathbf{V}_t = \gamma M_{t-1}\begin{bmatrix} \frac{1}{\alpha} & 1 \\ 1 & \alpha(1+\frac{\beta}{N\gamma}S_{t-1}(\alpha) \end{bmatrix} \tag{5.5}$$

The determinant of $\mathbf{V}_t$ is $\gamma M_{t-1}^2 \frac{\beta}{N} S_{t-1}(\alpha)$ and its inverse is

$$\mathbf{V}_t^{-1} = \frac{1}{\frac{\beta}{N}S_{t-1}(\alpha)M_{t-1}}\begin{bmatrix} \alpha(1+\frac{\beta}{N\gamma}S_{t-1}(\alpha)) & -1 \\ -1 & \frac{1}{\alpha} \end{bmatrix}. \tag{5.6}$$

5.1.2 Discrete Likelihood Function

Set $\mathbf{Z}_t = (Y_t, M_t)^t$. The joint density of $\mathbf{Z} = (\mathbf{Z}_1^t, \ldots, \mathbf{Z}_T^t)^t$ is a product of normal multivariate densities

$\prod_{t=1}^{T} f_N(\mathbf{Z}_t \,|\, \boldsymbol{\mu}_{t-1}, \mathbf{V}_{t-1})\,\mathbf{dZ}_t$ indexed by parameters α, β and γ . The likelihood function for α, β and γ given observed data is

$$\ln \mathcal{L}(\alpha,\beta,\gamma \,|\, \mathbf{Y}_T, \mathbf{M}_T) \propto -\frac{T}{2}(\ln\beta + \ln\gamma) - \frac{1}{2}\sum_{t=1}^{T}\ln S_t(\alpha) - \sum_{t=1}^{T}\ln I_t(\alpha) - \frac{1}{2}\sum_{t=1}^{T}Q_t(\alpha,\beta,\gamma) \tag{5.7}$$

In terms of $S_t(\alpha)$ and $I_t(\alpha) = \frac{M_t}{\alpha}$ the exponent of the log likelihood is (Appendix 5.2)

$$Q_t(\alpha,\beta,\gamma) = \frac{1}{\gamma I_t(\alpha)}(\Delta Y_t + \gamma I_{t-1}(\alpha))^2 + \frac{N}{\beta S_{t-1}(\alpha) I_{t-1}(\alpha)}(\Delta S_t(\alpha) + \frac{\beta}{N}S_{t-1}(\alpha)I_{t-1}(\alpha))^2 \tag{5.8}$$

At $\alpha = 1$ $Q_t(1,\beta,\gamma)$ is identical to the quadratic form appearing in the likelihood function in Eqn. (4.3). For α fixed and known, sufficient statistics for inference about β and γ are functionally identical to those defined in Section 4 so a conditional on α maximizers of β and γ satisfy the same conditions as in the case $\alpha = 1$. To facilitate computation expand

$$Q_t(\alpha,\beta,\gamma) = \frac{(\Delta Y_t)^2}{\gamma I_t(\alpha)} + 2\Delta Y_t + \frac{\gamma I_{t-1}(\alpha)}{2} + \frac{N(\Delta S_t(\alpha))^2}{\beta S_{t-1}(\alpha) I_{t-1}(\alpha)} + 2\Delta S_t(\alpha) + \frac{\beta}{N}S_{t-1}(\alpha)I_{t-1}(\alpha). \tag{5.9}$$

Remarks: Eqn. (5.9) differs in functional form from the likelihood function for γ and β for a single observational pair as shown in to Appendix A.4.3 in the appearance of two extra terms $2\Delta Y_t$ and $2\Delta S_t(\alpha)$. Only $2\Delta S_t(\alpha)$ depends on α. The likelihood function for γ and β in the absence of under/overcounting (Appendix A.4.3) is

$$-\frac{T}{2}\ln\gamma - \frac{1}{2\gamma}\sum_{t=1}^{T}\frac{(\Delta S_t + \Delta I_t)^2}{I_{t-1}} - \frac{\gamma}{2}\sum_{t=1}^{T} I_{t-1} \quad -\frac{T}{2}\ln\beta - \frac{N}{2\beta}\sum_{t=1}^{T}\frac{(\Delta S_t)^2}{S_t I_{t-1}} - \frac{\beta}{2N}\sum_{t=0}^{T} S_{t-1} I_{t-1}\,.$$

and in that case, sufficient statistics for inference about β and γ are, as in Eqn. (4.4), T and

$$Z_{1T}^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta S_t + \Delta I_t)^2}{I_{t-1}}\,,\quad Z_{2T}^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta S_t)^2}{S_{t-1} I_{t-1}}\,,\quad \bar{I}_T = \frac{1}{T}\sum_{t=1}^{T} I_{t-1} \text{ and } \bar{U}_T = \frac{1}{T}\sum_{t=1}^{T} I_{t-1} S_{t-1}\,.$$

The undercount quadratic form $Q_t(\alpha,\beta,\gamma)$ possesses functionally identical sufficient statistics along with two extra terms $\alpha\Delta Y_t$ and $\Delta S_t(\alpha)$. Define

$$Z_{1T}^2(\alpha) = \frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta Y_t)^2}{I_{t-1}(\alpha)}\,,\quad Z_{2T}^2(\alpha) = \frac{1}{T}\sum_{t=1}^{T}\frac{(\Delta S_t(\alpha))^2}{S_{t-1}(\alpha) I_{t-1}(\alpha)}\,, \tag{5.10a}$$

$$\bar{I}_T(\alpha) = \frac{1}{T}\sum_{t=1}^{T} I_{t-1}(\alpha) \text{ and } \bar{U}_T(\alpha) = \frac{1}{T}\sum_{t=1}^{T} I_{t-1}(\alpha) S_{t-1}(\alpha)\,. \tag{5.10b}$$

Assertion 5.1: The joint likelihood function for α, β and γ is

$$\ln\mathcal{L}(\alpha,\beta,\gamma\,|\,\mathbf{Y}_T,\mathbf{M}_T) \propto -\frac{T}{2}(\ln\beta + \ln\gamma) + T\ln\alpha - \frac{1}{2}\sum_{t=1}^{T}\ln S_t(\alpha) - \frac{1}{2}\sum_{t=1}^{T} Q_t(\alpha,\beta,\gamma) \text{ with}$$

$$\sum_{t=1}^{T} Q_t(\alpha,\beta,\gamma) = T\left[\frac{Z_1^2(\alpha)}{2\gamma} + \frac{\gamma}{2}\bar{I}(\alpha) + \frac{N}{2\beta}Z_{2T}^2(\alpha) + \frac{S_T(\alpha) - S_0(\alpha)}{T} + \frac{\beta}{2N}\bar{U}_T(\alpha)\right] \tag{5.11}$$

5.3 Undercount Maximum Likelihood Estimation

Compute MLEs of β and γ as functions of α using quadratic root formulae as in the absence of undercounting. Then identify a joint MLE of a, β and γ by computing the joint likelihood function of $\beta(\alpha)$ and $\gamma(\alpha)$ over an allowable range of α.

5.3.1 Monte Carlo Simulation of Undercount MLE Estimates

A Monte Carlo simulation of MLEs of α, β and γ show these estimates to be unbiased and to approximate true parameter values well. The first Monte Carlo study assumes sampling over $t = 0$ to 50 This sample interval spans the life cycle of infections, from initial rise, to a peak and then decline. Any robust estimation method must perform well with such complete samples. If undercounting is ignored MLEs are biased (Tables 5.2 and Figures 5.5 and 5.6). The sample size for the second study stops at $t = 20$—prior to median/mean peak times of 34. It yields approximately unbiased MLEs of all three parameters.

Table 5.1 Undercount MLE Estimates

$$N = 1000 \;\; I(0) = 30 \;\; \alpha = 0.7 \;\; \beta = 0.2 \;\; \gamma = 0.1$$
$$Time\ Periods = 50 \;\; Simulation\ Runs = 500$$

	Alpha	Beta	Gamma	Max Infected	Peak Time
Min	0.4919	0.1540	0.0702	121	21
First Q	0.6635	0.1924	0.0944	165	31
Median	**0.6999**	**0.2001**	**0.0998**	**183**	**34**
Mean	**0.7006**	**0.2001**	**0.1004**	**183**	**34**
Third Q	0.7354	0.2075	0.1064	198	38
Max	0.9258	0.2440	0.1328	264	49
Std Dev	0.0585	0.0125	0.0094	25.99	5.26

Figure 5.2 Alpha MLEs Boxplot

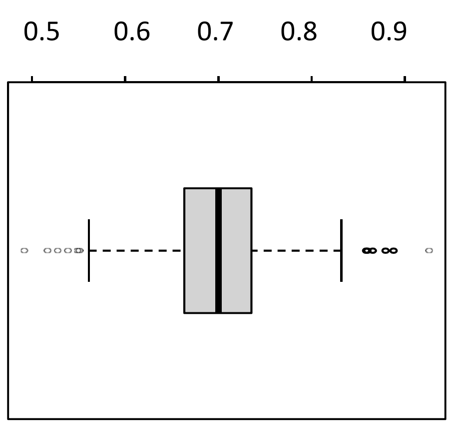


Figure 5.3 Monte Carlo Log Likelihood Samples

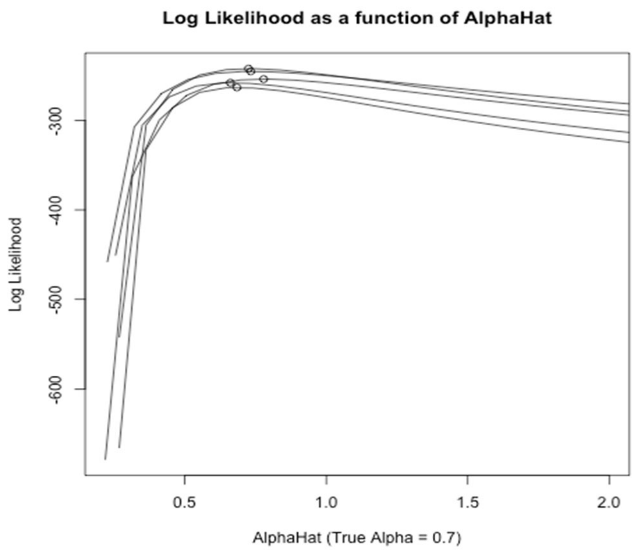
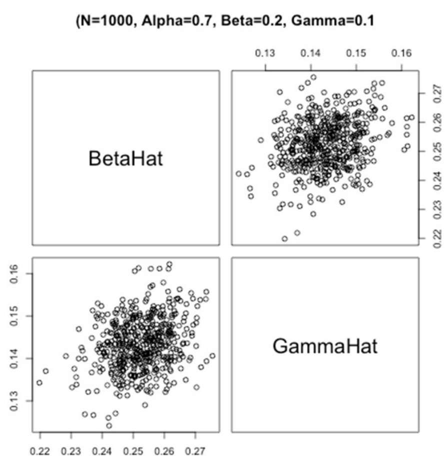


Figure 5.4 Undercount Beta-Gamma Scatterplots



A Q-Q plot of Alpha shows curvature away from normality in the tails, as do Q-Q plots of β and γ .

Figure 5.5 Alpha Q-Q Plot

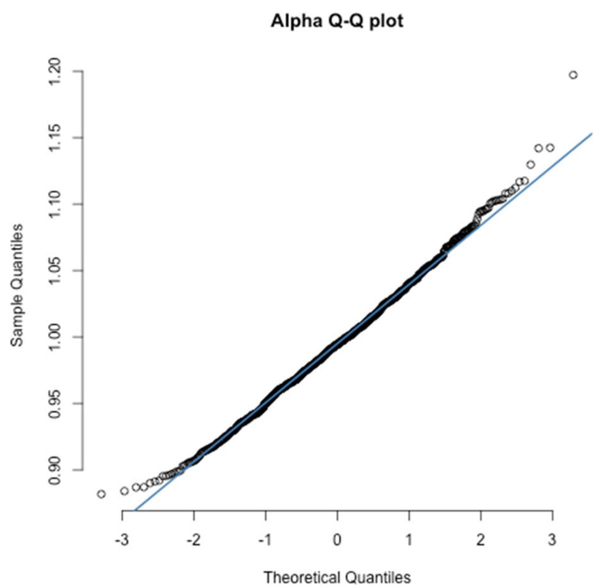


Table 5.2 displays MLEs of β and γ assuming no undercounting when in fact undercounting is present. Both estimates are, on average, substantially biased upward.

Table 5.2 Estimates ignoring undercounts

Parameter values $\beta = 0.2, \gamma = 0.1$ and α=0.7

	Beta	Gamma	Max Infected	Peak Time
Min	0.2289	0.1292	85	21
FirstQ	0.2461	0.1394	118	31
Median	**0.2520**	**0.1428**	**128**	**34**
Mean	**0.2523**	**0.1432**	**127**	**34**
ThirdQ	0.2587	0.1469	138	38
Max	0.2798	0.1604	172	49
StdDev	0.0093	0.0058	15.940	5.256

Figure 5.5 Biased Gamma

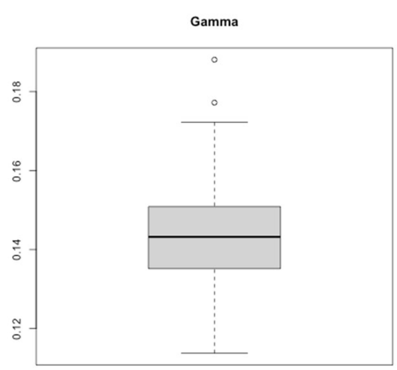
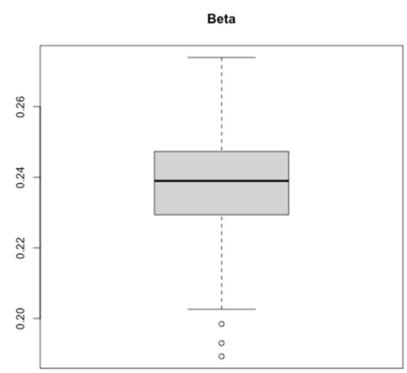


Figure 5.6 Biased Beta

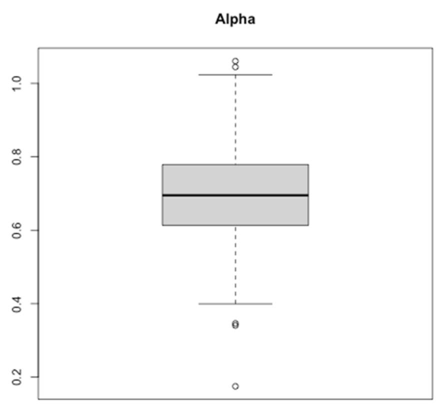


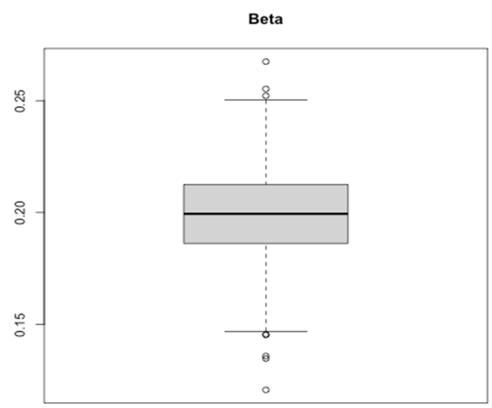
Monte Carlo results for a sample ranging over $t = 1,..20$ are displayed in Table 5.3 and Figures 5.7 and 5.8. MLEs appear to be unbiased.

Table 5.3 MLE Statistics for Undercount Sample Size 20

	Alpha	Beta	Gamma	Max Infected	Peak Time
Min	0.1749	0.1207	0.0261	50	15
First Q	0.6134	0.1862	0.0855	99	20
Median	**0.6952**	**0.1994**	**0.0998**	**122**	**20**
Mean	**0.6953**	**0.1992**	**0.0997**	**130**	**19.8**
ThirdQ	0.7786	0.2125	0.1133	153	20
Max	1.0607	0.2675	0.1640	651	20
StdDev	0.1285	0.0203	0.0201	49.068	0.687

Figure 5.7 Sample Size 20 **Figure 5.8 Sample Size 20**





APPENDICES

Sec. 4 MLE Proofs

A.4.1: The determinant of $\mathbf{\Sigma}_{t-1}$ is $\frac{\beta\gamma}{N} S_{t-1} I_{t-1}^2$. (A.1)

Proof:

$$\mathbf{\Sigma}_{t-1} = \frac{\beta}{N} S_{t-1} I_{t-1} \begin{bmatrix} 1 & -1 \\ -1 & 1+\frac{N\gamma}{\beta S_{t-1}} \end{bmatrix} = \begin{bmatrix} \frac{\beta}{N} S_{t-1} I_{t-1} & -\frac{\beta}{N} S_{t-1} I_{t-1} \\ -\frac{\beta}{N} S_{t-1} I_{t-1} & \frac{\beta}{N} S_{t-1} I_{t-1}(1+\frac{N\gamma}{\beta S_{t-1}}) \end{bmatrix} \text{so}$$

$$Det\left|\mathbf{\Sigma}_{t-1}\right| = (\frac{\beta}{N} S_{t-1} I_{t-1})^2 (1+\frac{N\gamma}{\beta S_{t-1}}) - (\frac{\beta}{N} S_{t-1} I_{t-1})^2 = (\frac{\beta}{N} S_{t-1} I_{t-1})^2 \times \frac{N\gamma}{\beta S_{t-1}} = \frac{\beta\gamma}{N} S_{t-1} I_{t-1}^2 \blacksquare$$

A.4.2: The inverse of $\mathbf{\Sigma}_{t-1}$ is $\frac{1}{\gamma I_{t-1}} \begin{bmatrix} (1+\frac{N\gamma}{\beta S_{t-1}}) & 1 \\ 1 & 1 \end{bmatrix}$. (A.2)

Proof:

$$\mathbf{\Sigma}_{t-1}^{-1}=\begin{bmatrix}\frac{\beta}{N}S_{t-1}I_{t-1} & -\frac{\beta}{N}S_{t-1}I_{t-1}\\ -\frac{\beta}{N}S_{t-1}I_{t-1} & \frac{\beta}{N}S_{t-1}I_{t-1}(1+\frac{N\gamma}{\beta S_{t-1}})\end{bmatrix}^{-1}=\frac{1}{Det|\mathbf{\Sigma}_{t-1}|}\times\begin{bmatrix}\frac{\beta}{N}S_{t-1}I_{t-1}(1+\frac{N\gamma}{\beta S_{t-1}}) & \frac{\beta}{N}S_{t-1}I_{t-1}\\ \frac{\beta}{N}S_{t-1}I_{t-1} & \frac{\beta}{N}S_{t-1}I_{t-1}\end{bmatrix}$$

So $\mathbf{\Sigma}_{t-1}^{-1}$

$$=\frac{1}{\frac{\beta\gamma}{N}S_{t-1}I_{t-1}^{2}}\times\begin{bmatrix}\frac{\beta}{N}S_{t-1}I_{t-1}(1+\frac{N\gamma}{\beta S_{t-1}}) & \frac{\beta}{N}S_{t-1}I_{t-1}\\ \frac{\beta}{N}S_{t-1}I_{t-1} & \frac{\beta}{N}S_{t-1}I_{t-1}\end{bmatrix}\begin{bmatrix}\frac{1}{\gamma I_{t-1}}(1+\frac{N\gamma}{\beta S_{t-1}}) & \frac{1}{\gamma I_{t-1}}\\ \frac{1}{\gamma I_{t-1}} & \frac{1}{\gamma I_{t-1}}\end{bmatrix}=\frac{1}{\gamma I_{t-1}}\begin{bmatrix}(1+\frac{N\gamma}{\beta S_{t-1}}) & 1\\ 1 & 1\end{bmatrix}\blacksquare$$

A.4.3: The likelihood function for γ and β is proportional to

$$-\frac{T}{2}\ln\gamma-\frac{1}{2\gamma}\sum_{t=1}^{T}\frac{(\Delta S_t+\Delta I_t)^2}{I_{t-1}}-\frac{\gamma}{2}\sum_{t=1}^{T}I_{t-1}\quad-\frac{T}{2}\ln\beta-\frac{N}{2\beta}\sum_{t=1}^{T}\frac{(\Delta S_t)^2}{S_tI_{t-1}}-\frac{\beta}{2N}\sum_{t=0}^{T}S_{t-1}I_{t-1}\,.$$

Proof: Substitute $|\mathbf{\Sigma}_{t-1}|=\frac{\beta\gamma}{N}S_{t-1}I_{t-1}^{2}$, $\mathbf{\Sigma}_{t-1}^{-1}=\frac{1}{\gamma I_{t-1}}\begin{bmatrix}1+\frac{\gamma N}{\beta S_{t-1}} & 1\\ 1 & 1\end{bmatrix}$ and

$\mathbf{y}_t-\mathbf{\mu}_{t-1}=(\Delta S_{t-1}+\frac{\beta}{N}S_{t-1}I_{t-1},\Delta I_{t-1}-\frac{\beta}{N}S_{t-1}I_{t-1}+\gamma I_{t-1})$ into Eqn. (3.3) and rewrite a generic term of the log likelihood $\sum_{t=1}^{T}\{\ln|\mathbf{\Sigma}_{t-1}|-\frac{1}{2}(\mathbf{y}_t-\mathbf{\mu}_{t-1})^t\mathbf{\Sigma}_{t-1}^{-1}(\mathbf{y}_t-\mathbf{\mu}_{t-1})\}$ as proportional to $-\frac{1}{2}\ln\beta-\frac{1}{2}\ln\gamma$

$$-\frac{1}{2\gamma I_{t-1}}(\Delta S_{t-1}+\frac{\beta}{N}S_{t-1}I_{t-1},\ \Delta I_{t-1}-\frac{\beta}{N}S_{t-1}I_{t-1}+\gamma I_{t-1})^t\begin{bmatrix}1+\frac{\gamma N}{\beta S_{t-1}} & 1\\ 1 & 1\end{bmatrix}(\Delta S_{t-1}+\frac{\beta}{N}S_{t-1}I_{t-1},\ \Delta I_{t-1}-\frac{\beta}{N}S_{t-1}I_{t-1}+\gamma I_{t-1})$$

$$=-\frac{1}{2}\ln\beta-\frac{1}{2}\ln\gamma-\frac{1}{2\gamma I_{t-1}}(\Delta S_{t-1}+\Delta I_{t-1}+\gamma I_{t-1})^2-\frac{N}{2\beta S_{t-1}I_{t-1}}(\Delta S_{t-1}-\frac{\beta}{N}S_{t-1}I_{t-1})^2\,. \qquad \text{(A.4)}$$

Factoring out I_{t-1} in the first quadratic form above and $S_{t-1}I_{t-1}$ in the second quadratic form above write (A.4) as

$$-\frac{1}{2}\ln\beta-\frac{I_{t-1}}{2\gamma}(\frac{\Delta S_{t-1}+\Delta I_{t-1}}{I_{t-1}}+\gamma)^2-\frac{N}{2\beta}S_{t-1}I_{t-1}(\frac{\Delta S_{t-1}}{S_{t-1}I_{t-1}}+\frac{\beta}{N})^2. \tag{A.5}$$

Expanding terms, the first quadratic form in (A.5) is $\frac{(\Delta S_{t-1}+\Delta I_{t-1})^2}{\gamma I_{t-1}}+2(\Delta S_{t-1}+\Delta I_{t-1})+\gamma I_{t-1}$ and the second is $-\frac{N}{2\beta}\frac{\Delta S_{t-1}^2}{S_{t-1}I_{t-1}}-\Delta S_{t-1}-\frac{\beta}{2N}S_{t-1}I_{t-1}$. In turn, the log likelihood is proportional to

$$-\frac{T}{2}\ln\gamma-\frac{1}{2\gamma}\sum_{t=1}^{T}\frac{(\Delta S_{t-1}+\Delta I_{t-1})^2}{I_{t-1}}-\frac{\gamma}{2}\sum_{t=1}^{T}I_{t-1}\quad-\frac{T}{2}\ln\beta-\frac{N}{2\beta}\sum_{t=1}^{T}\frac{\Delta S_{t-1}^2}{S_{t-1}I_{t-1}}-\frac{\beta}{2N}\sum_{t=1}^{T}S_{t-1}I_{t-1}.\ \blacksquare \tag{A.6}$$

In terms of statistics defined in Eqn. (4.4), (A.6) is *T* times

$$-\frac{1}{2}\ln\gamma-\frac{1}{2\gamma}Z_1^2-\frac{\gamma}{2}\bar{I}_T\quad-\frac{1}{2}\ln\beta-\frac{N}{2\beta}Z_2^2\ -\frac{\beta}{2N}\bar{U}_T\ \bullet \tag{A.7}$$

A.4.4 Necessary conditions for $\hat{\gamma}$ and $\hat{\beta}$ to be maximizers of (4.3) are

$$\frac{Z_{1T}^2}{\gamma^2}-\frac{1}{\gamma}-\bar{I}_T=0 \qquad \text{or} \qquad \bar{I}_T\gamma^2+\gamma-Z_{1T}^2=0 \tag{A.8}$$

and

$$\frac{N}{\beta^2}Z_{2T}^2-\frac{1}{\beta}-\frac{1}{N}\bar{U}_T=0 \quad \text{or} \quad -\beta+NZ_2^2-\frac{\beta^2}{N}\bar{U}_T=0\ . \tag{A.9}$$

For β and γ greater than zero, unique roots are

$$\frac{d\ln L}{d\gamma}=0 \Rightarrow \hat{\gamma}=\frac{-1+\sqrt{1+4\bar{I}_T Z_{1T}^2}}{2\bar{I}_T}. \tag{A. 10}$$

and

$$\frac{d\ln L}{d\beta}=0 \Rightarrow \hat{\beta}=N\times\frac{-1+\sqrt{1+4\bar{U}_T Z_{2T}^2}}{2\bar{U}_T}\ . \tag{A.11}$$

A.4.5 At $\boldsymbol{\theta}=(\hat{\gamma},\hat{\beta})^t$ the matrix $\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ln\ \mathcal{L}<\mathbf{0}$.

Proof: The inequality $\frac{d^2 \ln L}{d\gamma^2} < 0$ implies $-\frac{2Z_{1T}^2}{\gamma^3} + \frac{1}{\gamma^2} < 0$ or $\gamma - 2Z_{1T}^2 < 0.$ At $\gamma = \hat{\gamma} > 0$ the inequality $\hat{\gamma} < 2Z_{1T}^2$ is equivalent to $\frac{-1+\sqrt{1+4\bar{I}Z_{1T}^2}}{2\bar{I}_T} < 2Z_{1T}^2$ or $\sqrt{1+4\bar{I}Z_{1T}^2} < 1+4\bar{I}_T Z_{1T}^2$. Hence $\bar{I}_T Z_{1T}^2 > 0$ is sufficient.

The inequality $\frac{d^2 \ln L}{d\beta^2} < 0$ is $-\frac{2NZ_{2T}^2}{\beta^3} + \frac{1}{\beta^2} < 0$ or $\beta < 2NZ_{2T}^2$. At $\beta = \hat{\beta}$ the inequality $\beta < 2NZ_{2T}^2$ is $\frac{-1+\sqrt{1+4\bar{U}_T Z_{2T}^2}}{2\bar{U}_T} < 2NZ_{2T}^2$ or $\sqrt{1+4\bar{U}_T Z_{2T}^2} < 1+4N\bar{U}_T Z_{2T}^2$ which obtains for positive $\bar{U}_T Z_{2T}^2$. Hence at $\boldsymbol{\theta} = (\hat{\gamma}, \hat{\beta})^t$ the matrix $\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln \mathcal{L} < \mathbf{0}$. ■

Sec. 5 Undercount MLE Proofs

Assertion 6: The conditional mean and variance of I_t given Y_t, Y_{t-1} and I_{t-1} are

$$E(I_t | Y_t; Y_{t-1}, I_{t-1}) = E(I_t | Y_{t-1}, I_{t-1}) + \Delta Y_t = I_{t-1} + \frac{\beta}{N} I_{t-1}(Y_{t-1} - I_{t-1}) + \Delta Y_t$$

and

$$Var(I_t | Y_t; Y_{t-1}, I_{t-1}) = Var(I_t | Y_{t-1}, I_{t-1}) - \gamma I_{t-1} = \frac{\beta}{N}(Y_{t-1} - I_{t-1}) I_{t-1} = Var(S_t | S_{t-1}, I_{t-1}).$$

Proof:

The joint distribution of $(S_t, I_t)^t$ given S_{t-1} and I_{t-1} is bivariate normal with mean

$$\begin{pmatrix} E(S_t | S_{t-1}, I_{t-1}) \\ E(I_t | S_{t-1}, I_{t-1}) \end{pmatrix} = \begin{pmatrix} -\frac{\beta}{N} S_{t-1} I_{t-1} \\ \frac{\beta}{N} S_{t-1} I_{t-1} - \gamma I_{t-1} \end{pmatrix}$$

and

$$Var\left(\begin{pmatrix} S_t \\ I_t \end{pmatrix} | S_{t-1}, I_{t-1}\right) = \begin{bmatrix} \frac{\beta}{N} S_{t-1} I_{t-1} & -\frac{\beta}{N} S_{t-1} I_{t-1} \\ -\frac{\beta}{N} S_{t-1} I_{t-1} & \frac{\beta}{N} S_{t-1} I_{t-1} + \gamma I_{t-1} \end{bmatrix}.$$

When undercounting takes place $S_t + I_t - M_t = Y - M_t$ and M_t are observed, not S_t and I_t.

The mean value of $Y_t = S_t + I_t = E(S_t + I_t) = S_{t-1} + I_{t-1} - \gamma I_{t-1}$. Substituting $S_t = Y_{t-1} - I_{t-1}$

$$\begin{pmatrix} E(Y_t | Y_{t-1}, I_{t-1}) \\ E(I_t | Y_{t-1}, I_{t-1}) \end{pmatrix} = \begin{pmatrix} Y_{t-1} - \gamma I_{t-1} \\ \frac{\beta}{N} I_{t-1}(Y_{t-1} - I_{t-1}) + (1-\gamma) I_{t-1} \end{pmatrix}$$

Use

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & -a \\ -a & a+b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} b & b \\ b & a+b \end{bmatrix}$$

to arrive at

$$Var(\begin{pmatrix} Y_t \\ I_t \end{pmatrix} | S_{t-1}, I_{t-1}) = \begin{bmatrix} \gamma I_{t-1} & \gamma I_{t-1} \\ \gamma I_{t-1} & \frac{\beta}{N} S_{t-1} I_{t-1} + \gamma I_{t-1} \end{bmatrix} = \begin{bmatrix} \gamma I_{t-1} & \gamma I_{t-1} \\ \gamma I_{t-1} & \frac{\beta}{N}(Y_{t-1} - I_{t-1}) I_{t-1} + \gamma I_{t-1} \end{bmatrix}.$$

The variance of $I_t \text{ given } Y_t$ is

$$Var(I_t | Y_t; Y_{t-1}, I_{t-1}) = \frac{\beta}{N} S_{t-1} I_{t-1} + \gamma I_{t-1} - \gamma I_{t-1} = \frac{\beta}{N}(Y_{t-1} - I_{t-1}) I_{t-1}.$$

In turn the conditional expectation of $I_t \text{ given } Y_t$ is

$$E(I_t | Y_t; Y_{t-1}, I_{t-1}) = E(I_t | Y_{t-1}, I_{t-1}) + \frac{Cov(I_t, Y_t | Y_{t-1}, I_{t-1})}{Var(Y_t | Y_{t-1}, I_{t-1})} \; [Y_t - E(Y_t | Y_{t-1}, I_{t-1})].$$

Here $\frac{Cov(I_t, Y_t | Y_{t-1}, I_{t-1})}{Var(Y_t | Y_{t-1}, I_{t-1})} = 1$ so $E(I_t | Y_t; Y_{t-1}, I_{t-1}) = E(I_t | Y_{t-1}, I_{t-1}) + [Y_t - E(Y_t | Y_{t-1}, I_{t-1})]$.

Define $\Delta Y = Y_t - Y_{t-1} \; and \; \Delta I_t = I_t - I_{t-1}$. Substitution yields

$$E(I_t | Y_t; Y_{t-1}, I_{t-1}) = I_{t-1} + \Delta Y_t + \frac{\beta}{N} I_{t-1}(Y_{t-1} - I_{t-1}) \; \blacksquare$$

REFERENCES

Abadie, A. Bertolotti,p. Deamer. B..Sarler.A. and Shah, D. (2021) Epidemic Modeling and Estimation IDSS COVID-19 Collaboration (ISOLAT) Project, IDSS, MIT

Aït Sahalia ,Y. (2002) Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closedb Form Approximation Approach Econometrica Vol. 70 No. 1 223-262

Allen, L. (2017) *A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis* Inf. Dis. Modeling pp 128-142. Publ. online Mar 2017.

Lo, A.W. (1988) Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data *Econometric Theory*, Vol. 4, No. 2, pp. 231-247.

Sørensen, M. (1983) On Maximum Likelihood Estimation in Randomly Stopped Diffusion-Type Processes Int. Stat, Review Vol. 51, No. 1 pp. 93-110 Published by: International Statistical Institute (ISI) Stable URL: https://www.jstor.org/stable/1402735 April 1983

van den Driessche, Infect Dis Model. (2017) Aug; 2(3): 288–303.Published online 2017 Jun 29. doi: 10.1016/j.idm.2017.06.002