On the Effects of Type II Left Censoring in Stable and Chaotic Compartmental Models for Infectious Diseases: Do Small Sample Estimates Survive Censoring?

Alessandro Maria Selvitella[1, 3, 4], **Kathleen Lois Foster**[2, 4]

[1] Department of Mathematical Sciences, Purdue University Fort Wayne, Fort Wayne, Indiana, 46805, United States
[2] Department of Biology, Ball State University, Muncie, Indiana, 47306, United States
[3] eScience Institute, University of Washington, Seattle, WA 98195, United States
[4] NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, Illinois, IL 60208, United States
aselvite@pfw.edu, klfoster@bsu.edu

Abstract

In this paper, we discuss a selection of tools from dynamical systems and order statistics, which are most often utilized separately, and combine them into an algorithm to estimate the parameters of mathematical models for infectious diseases in the case of small sample sizes and left censoring, which is relevant in the case of rapidly evolving infectious diseases and remote populations. The proposed method relies on the analogy between survival functions and the dynamics of the susceptible compartment in SIR-type models, which are both monotone decreasing in time and are both determined by a dual variable: the hazard function in survival prediction and the number of infected people in SIR-type models. We illustrate the methodology in the case of a continuous model in the presence of noisy measurements with different distributions (Normal, Poisson, Negative Binomial) and in a discrete model, reminiscent of the Ricker map, which admits chaotic dynamics. This estimation procedure shows stable results in experiments based on a popular benchmark dataset for SIR-type models and small samples. This manuscript illustrates how classical theoretical statistical methods and dynamical systems can be merged in interesting ways to study problems ranging from more fundamental small sample situations to more complex infectious disease and survival models, with the potential that this tools can be applied in the presence of a large number of covariates and different types of censored data.

1 Introduction

Most often, modern methodology in survival prediction is concerned with right censoring, namely situations in which the researchers only know the lower bound on survival time (Yu et al. 2011; Kvamme and Borgan 2019). A typical situation is that of patients surviving beyond the time until which a doctor is monitoring their diseases. Furthermore, machine learning fits the best the purpose of generalization which takes advantage of the large flow of data and extensive computational power to prove results valid to unseen instances and a large body of work has been dedicated to the development of survival prediction algorithms (Wang, Li, and Reddy 2022). More recently and motivated by the COVID-19 pandemic, machine learning researchers have dedicated their attention to the development and implementation of new algorithms for high-dimensional epidemiological data that take advantage of a long history of mathematical modeling of infectious disease dynamics (Brauer, Castillo-Chavez, and Feng 2019; Schutt, Foster, and Selvitella 2021; Olson, Foster, and Selvitella 2022; Schutt, Foster, and Selvitella 2022). In particular, researchers have developed new methodologies that combine the predictive power of machine learning algorithms with interpretable models regulated by coupled systems of nonlinear differential equations. Such equations, often called *compartmental models*, describe the trajectory of the disease, by dividing the population in compartments and by describing the flow from one compartment to another a system of nonlinear differential equations (Qian, Alaa, and van der Schaar 2020; Vega, Flores, and Greiner 2022). Some independent attention has been given to the potential impact of classical theoretical statistics tools, such as order statistics, which, for decades, have been for decades a source of results in engineering, reliability theory, and survival analysis (Arnold, Balakrishnan, and Nagaraja 2008). Indeed, the combination of machine learning and differential equation tools can benefit from such a long tradition of researchers in statistical theory who have studied the theoretical underpinnings on time-to-event data and in particular the asymptotic behaviour of order statistics and maximum likelihood estimators (MLEs) for large sample sizes and censored data (Bhattacharya 1985; Casella and Berger 2002; Arnold, Balakrishnan, and Nagaraja 2008).

In this paper, we illustrate a circle of ideas to connect SIR models with survival analysis methods and a combination of these methodologies to describe what the potential effect is of left censoring in the case of small samples which is the traditional situation in epidemiological models and typical for closed (Menchhofer et al. 2021) or data-disadvantaged communities (Foster and Selvitella 2020, 2021d,b). For such a purpose, we provide a maximum likelihood estimator (MLE) for noisy infectious disease data and show some of the potential effects of censoring and of neglecting censoring in a small sample situation or for systems of differential and difference equations which admit chaotic motion. We adapt an asymptotic theorem for general MLEs of Type II censored data based on non-parametric properties of order statistics to the case in which the underlining deterministic dynamics is governed by SIR-type epidemic models. For il-

lustrations, we use a publicly available dataset collected during the flu epidemic in an English boys schoolboard (763 individuals) from January 22nd, 1978 (day 0) to February 4th, 1978 (day 13) (BMJ 1978). We illustrate then the effect of left censoring in the presence of noise in the data, with noise modeled with equidispersed and overdispersed distributions (Normal, Poisson, and Negative Binomial). We also discuss an example of a discrete toy-model for infectious disease dynamics based on the *Ricker map* (Ricker 1954; Lutscher 2019), which admits chaotic motion for certain ranges of parameters. For this model, we show how crucial it is to not neglect censored data, by illustrating an example in which the likelihood without censored data provides estimates of the parameters which give rise to a chaotic dynamics even if the starting point of the algorithm is nearby a stable trajectory. We also show how in such a case, the MLE which includes left censoring terms is more stable. Note that the problem of left censored data is interesting in infectious disease models because it is important to not neglect even the minimal information available early on in a rapidly growing epidemic; early measurements are the most noisy and missing important information in the first periods after the first infection can have important negative consequences on predictions and, in turn, on decisions concerning non-pharmaceutical interventions (Qian, Alaa, and van der Schaar 2020; Foster and Selvitella 2021a,c; Selvitella et al. 2021).

The remaining part of this manuscript is organized as follows. In Section 2, we give some theoretical results, including introductory facts on SIR models and order statistics and Theorem 1 that covers thelarge sample asymptotics case of the MLE of the parameters of stochastic processes with mean function determined by an SIR model with random noise in the case of Type II left censored data. Section 3 describes our experiments for not chaotic time-continuous SIR models, while Section 4 is dedicated to the effect of left censoring in models with chaotic dynamics, such as an extension of the *Ricker model*. We conclude with the discussion of our results in Section 5 and conclusions in Section 6.

2 Theory

The SIR Model. Consider the following SIR model.

$$\begin{cases} \dot{S}(t) = -\beta S(t) I(t) \\ \dot{I}(t) = \beta S(t) I(t) - \gamma I(t) \\ \dot{R}(t) = -\gamma I(t) \end{cases} . \tag{1}$$

Here $t \in [,+\infty)$, $S(t) \geq 0, I(t) \geq 0, R(t) \geq 0$, $S(t), I(t), R(t) \in \mathcal{C}^1([0,+\infty)])$, and $S(t)+I(t)+R(t) = N$. The compartments S, I, R represent susceptible, infectious, and recovered individuals, respectively. The parameter $\beta > 0$ represents the transmission rate, while the parameter $\gamma > 0$ represents the death rate. We will consider the random variables $Y(t) := I(t) + \epsilon(t)$, such that $E[Y(t)] = I(t)$ for every $t \in [0,+\infty)$ and $Var(Y(t)) = Var(\epsilon(t)) = \sigma^2$, with $\epsilon(t)$ independent and identically distributed. In other words, $Y(t)$ is a non-stationary stochastic process with mean function $I(t)$ and constant variance σ^2. The three parameters of the model are therefore β, γ, σ^2. To maintain conservation of the total population, we will adjust the susceptible terms accordingly: $X(t) := S(t) - \epsilon(t)$.

Order Statistics. Consider a random sample $-\infty < Y_1, \dots, Y_n < +\infty$ and the corresponding order statistics $-\infty < Y_{1:n} \leq \dots \leq Y_{n:n} < +\infty$. Here $Y_{s:n}$ represents the s-th smallest order statistics out of a sample of n elements. In this framework, s is not a random variable. If the results of an experiment are observed only after $Y_{s:n}$, the s-th data element, we obtain a Type II left censored sample. Suppose $Y_i \sim Y$ are independent for $i = 1, \dots, n$ with pdf $f(y|\theta)$ and cdf $F(y|\theta)$. The likelihood function is given by

$$\mathcal{L}(\theta|y) = \frac{n!}{s!} F(y_s|\theta)^s \prod_{k=s+1}^{n} f(y_k|\theta), \quad y_s < \dots < y_n.$$

Here s is fixed, while $Y_{s:n}$ is random. Similarly, in the case of Type II right censored sample, the likelihood function is given by

$$\mathcal{L}(\theta|y) = \frac{n!}{r!} \left(1 - F(y_r|\theta)\right)^r \prod_{k=1}^{n-r} f(y_k|\theta), \quad y_1 < \dots < y_r.$$

Here r is fixed, while $Y_{r:n}$ is random. For large sample sizes, we have the following asymptotic theorem.

Theorem 1 *Consider the stochastic processes $X(t) = S(t) - \epsilon(t)$ and $Y(t) = I(t) + \epsilon(t)$ with $\epsilon(t) \sim WN(0, \sigma^2)$ (white noise - independent identically distributed random variables), and with $S(t), I(t)$ solving system* (1)*. Consider also a random sample of size n: $(X_1, Y_1), \dots, (X_n, Y_n)$ with $(X_i, Y_i) := (X(t_i), Y(t_i))$ for $i = 1, \dots, n$ and $t_1, \dots, t_n \in \mathbb{R}^+$ and the regularity assumptions of (Bhattacharya 1985) for the density of $\epsilon(t)$. Then*

1. *$P(X_{t+1} < X_t) \to 1$ as $\sigma^2 \to 0$.*
2. *The asymptotic distribution of the MLE $\hat{\theta} := (\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2)^T$ given by*

$$\hat{\theta} = argmin_{\theta \in \mathbb{R}^3_+} \mathcal{L}(\theta|\mathbf{X}; s) = \tag{2}$$

$$\frac{n!}{s!} F(y_s|\theta; t, S, R)^s \prod_{k=s+1}^{n} f(y_k|\theta; t, S, R), \tag{3}$$

$$y_s < \dots < y_n, \tag{4}$$

with $\mathbf{X} = [t_i; S_i, I_i, R_i]_{i=1}^n$ in the limit for $n \to +\infty$ is given by $\sqrt{n}(\hat{\theta} - \theta_M) \sim N(0, \mathbf{J^{-1}}(\theta_\mathbf{M}))$, with $\theta_M := (\beta_M, \gamma_M, \sigma_M^2)^T$ being the vector of the true values of the parameters and $\mathbf{J}(\theta_\mathbf{M})$ being the Fisher Information Matrix of θ_M.

Remark 1 *Note that the exact shape of the dependency $y|\theta; t, S, R$ is quite involved as it depends on the solution of equation* (1)*. The asymptotic theorem does not depend on the exact shape of the information matrix $J(\theta)$.*

Sketch of the Proof of Theorem 1. Since $X_n = S_n + \epsilon_n$, then $X_n - X_{n+1} = \epsilon_n + S_n - S_{n+1} - \epsilon_{n+1} \sim WN(S_n - S_{n+1}, 2\sigma^2)$. Define $\delta_n := \epsilon_n - \epsilon_{n+1}$. Then $P(|\delta_n - E(\delta_n)| \geq t) \leq 2\sigma^2/t^2 \to 0$ and so $P(X_{t+1} < X_t) \to 1$ as $\sigma^2 \to 0$. For the ease of notation, in this sketch

s	-Log Lik.	R_0	σ
0	60.73	4.13	18.52
1	57.50	4.13	19.16
2	53.98	4.12	19.67
3	50.67	4.09	20.49
4	44.36	3.84	17.64
5	37.52	3.34	14.04
6	34.33	3.18	15.01
7	30.41	3.01	15.58

Table 1: Normal Error - Algorithm 1a.

of the proof, we consider $dim(\theta) = 1$ (the general case is analogous). Following a similar argument from (Bhattacharya 1985): By Taylor expansion of the score equation $\sqrt{n}(\hat{\theta} - \theta_M) \simeq -L'(\theta_M|y)/L''(\theta_M|y)$, and so $\sqrt{n}(\hat{\theta} - \theta_M) \simeq n^{1/2} * L'(\theta_M|y)/J(\theta_0)$. On the other side $E[n^{1/2} * L'(\theta_M|y)/J(\theta_M)] = 0$ and $Var[n^{1/2} * L'(\theta_0|y)/J(\theta_M)] = 1/J(\theta_M)$. The result follows from Lindeberg-Feller central limit theorem (Varadhan 2021; Casella and Berger 2002). The complete proof follows closely ideas from (Arnold, Balakrishnan, and Nagaraja 2008; Bhattacharya 1985; Lawless 1982; Casella and Berger 2002). In the case of large samples and censoring of orders $r, s = o(n)$ (not just bounded in n as presented in Theorem 1), the effects of the censoring is relatively negligible as well with a similar argument.

Algorithm 1: MLE for SIR models with Type II left censored data

Input: $\mathbf{X} = [t_i; S_i, I_i, R_i]_{i=1}^n$ dataset; $\theta_0 = (\beta_0, \gamma_0, \sigma_0^2)$ starting values; s censored position.
1. Simulate SIR-model(1) with $\theta_n = (\beta_n, \gamma_n, \sigma_n^2)^T$.
2.a Compute $\mathcal{L}(\theta_n|\mathbf{X}; s)$ in eq. (2). **[Algorithm 1a]**
2.b Compute $\mathcal{L}(\theta_n|\mathbf{X}_{s+1}^n; s)$ in eq. (2). **[Algorithm 1b]**
Output: MLE $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2)^T$; $\hat{R}_0$ estimated basic reproduction number; Confidence bands for $I(t)$.

Theorem 1 covers the case of large sample asymptotics. From now on, we will concentrate on small sample sizes only, as the distribution theory in such cases is more involved. In particular, distributional assumptions on the parent distribution are required. In small sample problems, it is important to try to save as much as possible of what we know, as data might not speak for itself loud enough in small sample sizes.

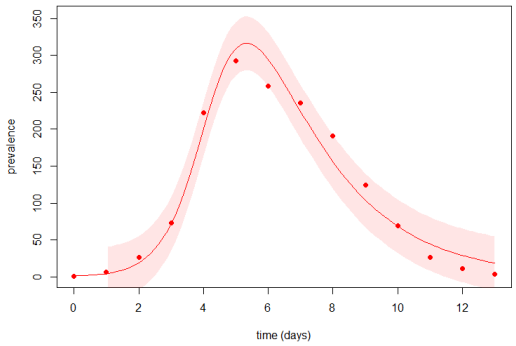
3 Experiments I - Stable Dynamics

In our experiments, we used Nelder-Mead algorithm (Nelder and Mead 1965) to optimize the likelihood in equation (2).

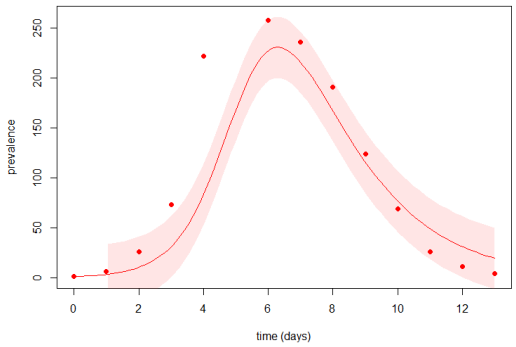
Normal Error Experiment

We fit the SIR model in equation (1) using both MLE Algorithm 1 (Algorithm 1 with likelihood 2.a) and Algorithm 1b (Algorithm 1 with likelihood 2.b) with a normally distributed error term. We start both algorithms with $\beta = \beta_0 = 0.003$, $\gamma = \gamma_0 = 0.48$, and $\sigma = \sigma_0 = 19$. Confidence bands are reported in Figure 1 and statistics in Tables 1 and 2.

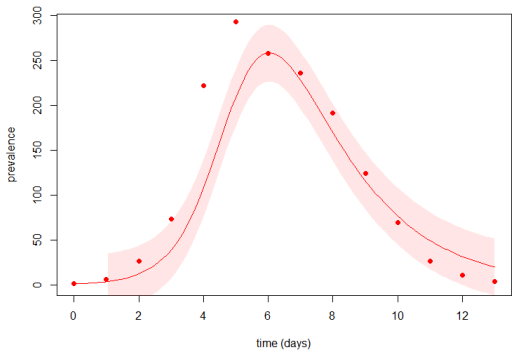
Normal Error - Algorithms 1a and 1b.



(a) Normal Error & MLE - Algorithm 1a & $s = 0$



(b) Normal Error & MLE - Algorithm 1a & $s = 7$



(c) Normal Error & MLE - Algorithm 1b & $s = 7$

Figure 1: Curve fit with 95 % Confidence Bands from Algorithm 1a based on Type II censored sample and Normal Error. a: $s = 0$. b: 1a & $s = 7$. c: 1b & $s = 7$.

s	-Log Lik.	R_0	σ
0	60.73	4.13	18.52
1	56.87	4.13	19.22
2	52.91	4.12	19.89
3	48.98	4.12	20.77
4	43.19	3.86	18.18
5	36.77	3.35	14.39
6	33.11	3.47	15.17
7	29.38	3.33	16.09

Table 2: Normal Error - Algorithm 1b

s	-Log Lik.	R_0	σ
0	70.73	4.08	10.03
1	68.84	4.07	20.05
2	65.24	4.04	19.70
3	62.78	4.01	19.48
4	55.52	3.74	17.65
5	49.27	3.33	16.54
6	45.94	3.18	16.19
7	40.61	3.03	19.41

Table 3: Poisson - Algorithm 1a

As expected, Algorithm 1a seems to decrease its performance with respect to Algorithm 1b after the time corresponding to the peak of $I(t)$ which happens at $t = 5$. After that time, the likehood $\mathcal{L}(\theta|\mathbf{X}; s)$ is not valid anymore, as we loose monotonicity. From there on, Algorithm 1a decreases performance and Algorithm 1b is more competitive. The decrease in performance can be even a potential criterion to detect the maximum number of infections.

Poisson vs Negative Binomial Errors Experiment

We fit the SIR model in equation (1) using MLE Algorithm 1a with Poisson distributed and Negative Binomial distributed error term. Algorithm 1a with error following a Poisson (equidispersed distribution) error starts with $\beta = \beta_0 = 0.003$, $\gamma = \gamma_0 = 0.48$, and $\sigma = \sigma_0 = 10$. Algorithm 1a with error following a Negative Binomial (overdispersed distribution, so higher sigma) error starts with $\beta = \beta_0 = 0.003$, $\gamma = \gamma_0 = 0.48$, $\sigma = \sigma_0 = 20$.

The model with a Poisson error has negative log-likelihood higher than both those with Normal and Negative Binomial error terms for all $s = 0, \ldots, 7$. Confidence bands are reported in Figure 2 and statistics in Tables 3 and 4.

s	-Log Lik.	R_0	σ
0	60.15	3.91	36.07
1	58.18	3.87	39.28
2	54.81	3.73	38.73
3	51.07	3.59	34.48
4	45.33	3.29	30.92
5	39.60	3.01	29.18
6	34.46	2.82	24.98
7	28.96	2.66	52.98

Table 4: Negative Binomial - Algorithm 1a

4 Experiments II - Chaotic Dynamics

Consider again the SIR system (1). If we solve the $S(t)$-equation, we obtain:

$$S(t) = S_0 \exp\left\{-\beta \int_0^t I(s)ds\right\}.$$

If we plug the solution of the $S(t)$-equation as a function of the variable $I(t)$ in the $I(t)$-equation, we obtain:

$$\dot{I}(t) = I(t)\left(\beta S_0 \exp\left\{-\beta \int_0^t I(s)ds\right\} - \gamma\right).$$

If we rescale the time t with respect to death rate units so that $\tau = t/\gamma$ and we intend the time derivative with respect to the scaled variable τ, we get

$$\dot{I}(\tau) = I(\tau)\left(R_0 \exp\left\{-\beta \int_0^{\tau\gamma} I(s)ds\right\} - 1\right), \quad (5)$$

with $R_0 := \beta S_0/\gamma$, the *basic reproduction number*. If we pass to the discrete dynamics $\dot{I}(\tau) \mapsto I_{\tau+1} - I_\tau$, we obtain:

$$I_{\tau+1} = I_\tau * R_0 \exp\left\{-\beta \sum_{s=0}^{[\tau\gamma]} I_s\right\}, \quad (6)$$

with $[x]$ defining the integer part of x. This system is very reminiscent of the *Ricker model* (Ricker 1954; Lutscher 2019), given by

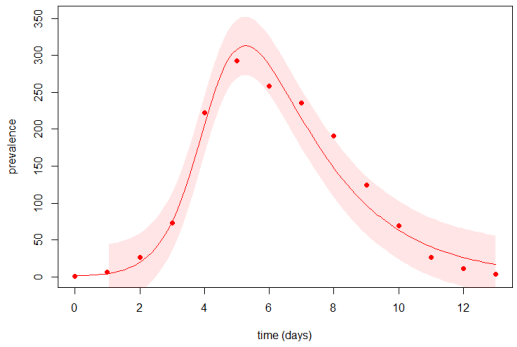
$$I_{\tau+1} = I_\tau e^{r\left(1-\frac{I_\tau}{k}\right)},$$

with $R_0 = \exp(r)$ (r is interpreted as an intrinsic growth rate of a population), $\beta = r/k$ (k is interpreted as the carrying capacity of the environment), and with I_τ being the expected number of individuals in generation τ. In the case of constant parameters, it is know that the SIR model is not chaotic (Brauer, Castillo-Chavez, and Feng 2019; Diekmann, Heesterbeek, and Britton 2013), while the *Ricker model* admits chaotic solutions for r in an appropriate range (Lutscher 2019). On the other side, it is well known that in the case of periodic transmission $\beta = \beta(t)$ with appropriate frequency, the SIR model presents chaotic dynamics (Brauer, van den Driessche, and Wu 2008; Earn, Rohani, and Grenfell 1998; Earn et al. 1998). Such a periodic term $\beta(t)$ morally re-weights the terms of the cumulative numbers of infectious diseases present in (5) and prevents solutions to decay to zero, so that its effect is that of resembling a Ricker-type dynamics. Therefore, we use the Ricker model as a toy model to illustrate the effect of left censoring in chaotic epidemiological models. Note that if we send $\tau \to +\infty$ on both sides of equation (6) and define $I_\infty := \lim_{\tau\to+\infty} I_\tau$, $I_{total} = \sum_{s=0}^{\infty} I_s$, we obtain

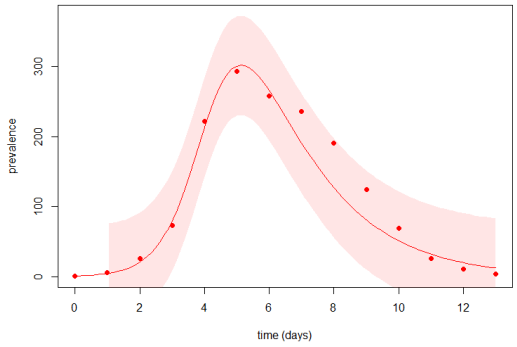
$$I_\infty = I_\infty * R_0 \exp\{-\beta I_{total}\},$$

which has a solution either $I_\infty = 0$ or $I_{total} = \frac{\ln R_0}{\beta}$ (in the case $I_{total} < +\infty$). The first solution is present also in the Ricker model and is an unstable equilibirum, while the second is reminiscent of the second fixed point of the

Poisson vs Negative Binomial Error - Algorithm 1a.



(a) Poisson — MLE - Algorithm 1a & $s = 0$



(b) Negative Binomial - Algorithm 1a & $s = 0$

Figure 2: Curve fit with 95 % Confidence Bands from Algorithm 1b based on Type II censored sample $s = 0$. (a): Poisson and (b): Negative Binomial Errors.

Ricker map, the one which goes from asymptotically stable for $0 < \ln R_0 < 2$ to chaotic for $\ln R_0$ closer and closer to 3 (Lutscher 2019). In the case in which the dynamics is chaotic, we cannot assume that the number of infected I_τ is monotone until its maximum I_{max} and then decreases. In fact, one expects many local maxima to appear in an a-periodic manner (Brauer, van den Driessche, and Wu 2008). A way to go is to couple the $I(t)$-equation with a $S(t)$-equation, estimate first the $S(t)$-equation, and then deduce the corresponding estimates for the $I(t)$-equation. This would require switching from left to right censoring in likelihood 2 and Algorithms 1a and 1b. We study the system

$$\begin{cases} S_{t+1} = S_t - \beta S_t I_t \\ I_{t+1} = I_\tau R_0 \exp\{-\beta I_t\} \\ R_t = N - I_t - S_t \end{cases} \tag{7}$$

$I_t \geq 0, S_t \geq 0, R_t \geq 0$ and $S_t + I_t + R_t = N$ for every $t \in \mathbb{N}$, with $N \in \mathbb{N}$ representing the total population size.

We can use Algorithm 1a and Algorithm 1b to illustrate some phenomena emerging in the presence of left censored data. We simulate system (7) with a population of

s	-Log Lik.	R_0	σ
0	862022	0.96	2.99
5	1882933	1.14	2.47
10	X	X	X
s	-Log Lik.	R_0	σ
0	862022	0.96	2.99
5	1818663	1.11	2.47
10	540963	12.26	15.45

Table 5: Ricker Model - Algorithms 1a (Up) and 1b (down).

$N = 1000$, parameters $R_0 = 1.5$, $\beta = 0.002$ and with an additive noise ϵ added to I_t distributed as $\epsilon \sim N(0, \sigma^2)$ with $\sigma^2 = 1$. We want to estimate these parameters, and we use $R_0 = 3$, $\beta = 1$, and $\sigma^2 = 9$ as starting values for Algorithm 1a vs Algorithm 1b for the $S(t)$ - equation, so using the right censored version of likelihood 2 (the number of individuals in the susceptible compartments is decreasing in time). We followed the dynamics until $S_t > 0.01$ using the full simulated data and right censoring for $s = 0, 5, 10$. The algorithm suffers very soon the complicated dynamics and even without censoring $s = 0$ the estimation behaves weirdly and looses the monotonicity of the S variable (See Figure 3). An interesting phenomenon appears with Algorithm 1b with censoring at $s = 10$ (See Figure 3). The starting points of the algorithm would provide stable solutions, however the noise takes over and the algorithm transitions quickly to a chaotic behaviour with a basic reproduction number $R_0 \simeq 12.16$, a value of the parameter where the Ricker dynamics for I is chaotic ($\ln 12.16 > 3$; Table 5). The use of Algorithm 1a does not show such a serious instability.

5 Discussion

We have gone through some simple examples and illustrate some potential problems in the analysis of infectious disease dynamics with survival analysis methods to motivate research on left censoring at the intersection of dynamical systems, theoretical statistics, and survival prediction. We believe a lot of great ideas might emerge with cross-pollination between these areas, also for high-dimensional models and large datasets. We believe that small sample estimates are still important and can survive censoring, especially in the case the dynamics is not-chaotic, once appropriate methodology is developed. Our goal has not been to optimize models and algorithms for predictive purposes, but to illustrate and review some of the problems that we might encounter with left censored data in infectious disease models and to promote a combination of tools that are usually not developed jointly. We also discussed several features of censored data methodology combined with ideas from survival analysis and differential equations to solve problems which can be encountered in the estimation of parameters in the case of small samples and potentially chaotic dynamics in infectious disease models. We connected the infectious disease dynamics of an SIR model to survival analysis by noticing that the dynamics of the susceptible compartment $S(t)$ has analogous functional properties of those of a survival function. Theorem 1 covers the case of large samples

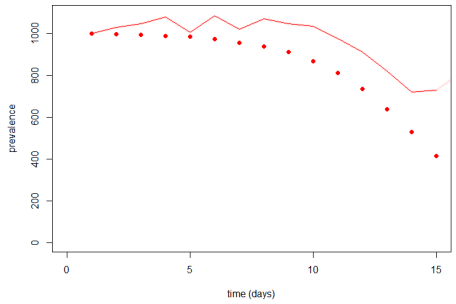
while the experiments the case of small sample size, which is a situation typically encountered in close environments, remote locations, or historical epidemics (Menchhofer et al. 2021; Selvitella et al. 2021; Earn et al. 2020). Given that our experiments are concerned primarily with the case of small data, the models cannot be too complicated and the data is not rich enough for generalization purposes.

The idea to enlarge the parameter space with a second equation in the Ricker model (7), adding the "susceptible equation" to the "infectious equation", in order to facilitate parameter estimation using survival prediction ideas and monotonicity, reminds of the idea of adding a momentum equation in stochastic gradient descent (SGD) (Nesterov 2013). GD might get stuck into local minima of non-convex functions and might require a large number of iterations to reach convergence because of noisy, oscillating minimizing sequences. The momentum equation aims at mitigating these two problems. In a similar way, adding the "susceptible equation" to the "infectious equation" in the case of chaotic motion straightens down the dynamics, makes convergence monotone, and does all of this without sacrificing biological interpretability as the $S(t)$)-equation represents the equation for the susceptible compartment. Note that the monotonocity problem of the $I(t)$-equation is present also in non-chaotic dynamics as $I(t)$ is never monotone in interesting cases ($R_0 > 1$) (Brauer, Castillo-Chavez, and Feng 2019), but the problem is more limited in the sense that the number of bumps has a more predictable distribution and the monotonicity of the censoring is valid in regular time intervals. Although Algorithm 1a has proven to be more stable for chaotic dynamics, in the presence of a noise term, than Algorithm 1b, the estimation of parameters in chaotic models remains very delicate as the ranges of parameters that give rise to non-chaotic solutions often occupy a restricted region of the parameter space. The idea of using survival prediction methods in combination with differential equation models is underdeveloped but not new in the literature. In (KhudaBukhsh et al. 2021), the authors show that solutions to coupled systems of nonlinear ordinary differential equations describing the large-population limits of Markovian stochastic processes for epidemic models have a natural interpretation as survival functions or cumulative hazard functions. In (Tang et al. 2022), the authors introduce an ODE notion for survival analysis to provide a unified modeling framework and to enable the development of versatile and scalable procedures for statistical inference and parameter estimation. Note that data in survival models or infectious disease models might be censored in many different ways. Types of censoring include right/left/interval censoring or Type I/Type II/hybrid censoring and they all have their own specificity (Arnold, Balakrishnan, and Nagaraja 2008). For illustrative purposes, we treated only Type II censoring and left censoring. Left censoring is much less studied in survival prediction literature with respect to right censoring.

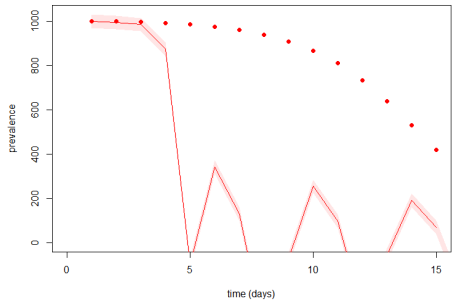
6 Conclusions

We illustrated a methodology to estimate the number of infected individuals in SIR-type models for small sample sizes in the case of a left censored data. We extended a large sample asymptotic theorem for MLEs (Arnold, Balakrishnan, and Nagaraja 2008) for left censored data to the case of noisy infectious disease trajectories. We ran multiple experiments using a popular small sample size dataset for testing the fit of SIR models (BMJ 1978). We discussed the appropriateness of the inclusion of left censoring information in Algorithms 1a vs 1b in the presence of Normal, Poisson, and Negative Binomial noise. We showed that including information about censored data in the likelihood is crucial for the stability of the estimates, especially in models which allow chaotic motion. Further work needs to be done in the case of more complicated mathematical models for infectious disease dynamics and with different types of censoring (e.g. Type I censoring, random censoring, hybrid censoring). Although Theorem 1 suggests a simplified situation in the case of large sample size, under which regime the censoring of a finite number of observations is not a big problem, key is to address the high-dimensional case.

Ricker Model - Algorithm 1a vs Algorithm 1b



(a) Ricker Model - Algorithm 1a



(b) Ricker Model - Algorithm 1b

Figure 3: Curve fit with 95 % Confidence Bands from Algorithms 1a vs 1b based on Type II censored sample $s = 0$. The vertical axis is $X(t) = S(t) - \epsilon(t)$, while the horizontal axis is the time t. (a): Algorithm 1a. (b): Algorithm 1b.

Acknowledgements

AMS and KLF are supported by the collaborative NSF Awards # 2152789 and # 2152792 on *RUI: Collaborative Research: DMS/NIGMS 1: The mathematical laws of morphology and biomechanics through ontogeny*.

References

Arnold, B.; Balakrishnan, N.; and Nagaraja, H., eds. 2008. *A first course in order statistics*. Philadelphia (PA): Society of Industrial and Applied Mathematics.

Bhattacharya, G. 1985. The asymptotics of maximum likelihood and related estimators based on Type II censored data. *Journal of the American Statistical Association*, 80: 398–404.

BMJ. 1978. News and Notes. Epidemiology-Influenza in a boarding school. British Medical Journal:1.

Brauer, F.; Castillo-Chavez, C.; and Feng, Z., eds. 2019. *Mathematical Models in Epidemiology*. New York (NY): Springer.

Brauer, F.; van den Driessche; and Wu, J., eds. 2008. *Mathematical Epidemiology — Lecture notes in mathematics - Mathematical biosciences subseries*. Heidelberg (Germany): Springer.

Casella, G.; and Berger, R., eds. 2002. *Statistical Inference - 2nd edition*. Belmont (CA): Brooks/Cole Cengage Learning.

Diekmann, O.; Heesterbeek, H.; and Britton, T., eds. 2013. *Mathematical tools for understanding infectious disease dynamics*. Princeton (NJ): Princeton University Press.

Earn, D.; Ma, J.; Poinar, H.; Dushoff, J.; and Bolker, B. 2020. Acceleration of plague outbreaks in the second pandemic. *Proceedings of the National Academy of Sciences of the U.S.A.*, 117(44): 27703–27711.

Earn, D.; Rohani, P.; Bolker, B.; and Grenfell, B. 1998. A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453): 667–670.

Earn, D.; Rohani, P.; and Grenfell, B. 1998. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society of London, Series B*, 265(1390): 7–10.

Foster, K.; and Selvitella, . A. 2020. Societal and economic factors associated with COVID-19 indicate that developing countries suffer the most. *Technium Social Sciences Journal*, 10: 637–644.

Foster, K.; and Selvitella, A. 2021a. Bayesian detection and uncertainty quantification of the first change point of the COVID-19 case curve in the Midwest: Timeliness of non-pharmaceutical interventions. *International Conference on Learning Representations 2021 Workshop on AI for Public Health & International Conference on Learning Representations 2021 Workshop on Machine Learning for Preventing and Combating Pandemics. May 7th, 2021*.

Foster, K.; and Selvitella, A. 2021b. Government measures against COVID-19 must be determined according to the socio-economic status of the country. *International Conference on Learning Representations 2021 Workshop on AI for Public Health. May 7th, 2021*.

Foster, K.; and Selvitella, A. 2021c. A higher order Taylor expansion of the initial trajectory of COVID-19 cases and deaths via Bayesian hierarchical models: a toy problem and possible public health insights. *International Conference on Learning Representations 2021 Workshop on AI for Public Health. May 7th, 2021*.

Foster, K.; and Selvitella, A. 2021d. On the relationship between COVID-19 reported fatalities early in the pandemic and national socio-economic status predating the pandemic. *AIMS Public Health*, 8(4): 439–455.

KhudaBukhsh, W.; Choi, B.; Kenah, E.; and Rempała, G. 2021. Continuous and discrete-time survival prediction with neural networks. *Interface Focus*, 10: 20190048.

Kvamme, H.; and Borgan, O. 2019. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Lifetime Data Analysis*, 27: 710–736.

Lawless, J., ed. 1982. *Statistical Models & Methods for Lifetime Data*. New York (NY): Wiley.

Lutscher, F., ed. 2019. *Integrodifference equations in spatial ecology*. Cham (Switzerland): Springer.

Menchhofer, K.; Mills, N.; Foster, K.; and Selvitella, A. 2021. COVID-19 incidence in the Indiana's secondary school system through a conditional Gaussian model and an age-structured compartmental model. *International Conference on Learning Representations 2021 Workshop on Machine Learning for Preventing and Combating Pandemics. May 7th, 2021*.

Nelder, J.; and Mead, R. 1965. A simplex method for function minimization. *Computer Journal*, 7(4): 308–313.

Nesterov, Y., ed. 2013. *Introductory lectures on convex optimization : A basic course*. New York (NY): Springer.

Olson, N.; Foster, K.; and Selvitella, A. 2022. On the possibility of mode-collapse phenomena in combined machine learning and differential equation models for infectious diseases. *ICML 2022 The 1st Workshop on Healthcare AI and COVID-19*.

Qian, Z.; Alaa, A.; and van der Schaar, M. 2020. When and How to Lift the Lockdown? Global COVID-19 Scenario Analysis and Policy Assessment using Compartmental Gaussian Processes. In *Proceedings of the Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 10729–10740.

Ricker, W. 1954. Stock and Recruitment. *Journal of the Fisheries Research Board of Canada*, 11: 559–623.

Schutt, N.; Foster, K.; and Selvitella, A. 2021. On learning the effects of healthcare overextension on increased mortality rate in the COVID-19 pandemic. *International Joint Conference on Artificial Intelligence 2021 Workshop on AI for Social Good*.

Schutt, N.; Foster, K.; and Selvitella, A. 2022. An interpretable time-series model for predicting nurse shortages and planning optimal nurse scheduling and staffing during the COVID-19 pandemic. *ICML 2022 The 1st Workshop on Healthcare AI and COVID-19*.

Selvitella, A.; Carolan, L.; Smethers, J.; Hernandez, C.; and Foster, K. 2021. A spatio-temporal investigation of the growth rate of COVID-19 incidents in Ohio, USA, early in the pandemic. *The Ohio Journal of Science*, 121(2): 33–47.

Tang, K., W.a nd He; Xu, G.; ; and Zhu, J. 2022. Survival Analysis via Ordinary Differential Equations. *Journal of the American Statistical Association: Theory and Methods*, To appear.

Varadhan, S., ed. 2021. *Probability Theory*. Providence (RI): American Mathematical Society.

Vega, R.; Flores, L.; and Greiner, R. 2022. SIMLR: Machine Learning inside the SIR model for COVID-19 Forecasting. *Forecasting*, 4(1): 72–94.

Wang, P.; Li, Y.; and Reddy, C. 2022. Machine Learning for Survival Analysis: A Survey. *ACM Computing Surveys*, 51(6): 1–36.

Yu, C.-N.; Greiner, R.; Lin, H.-C.; and Baracos, V. 2011. Learning Patient-Specific Cancer Survival Distributions as a Sequence of Dependent Regressors. In *Proceedings of the Advances in Neural Information Processing Systems 24 (NIPS 2011)*.