Fitting coefficients of differential systems with Monte Carlo methods

Christian Chan Shio[*] — Francine Diener[**]

[*] Mathematics Department
Ateneo de Manila University
Quezon City, Philippines
cchanshio@ateneo.edu

[**] Laboratoire J. A. Dieudonné
Université de Nice-Sophia Antipolis
CNRS, UMR 7351
06100 Nice, France
diener@unice.fr

ABSTRACT. We consider the problem of estimating the coefficients in a system of differential equations when a trajectory of the system is known at a set of times. To do this, we use a simple Monte Carlo sampling method, known as the rejection sampling algorithm. Unlike deterministic methods, it does not provide a point estimate of the coefficients directly, but rather a collection of values that "fits" the known data well. An examination of the properties of the method allows us not only to better understand how to choose the different parameters when implementing the method, but also to introduce a more efficient method by using a new two-step approach which we call sequential rejection sampling. Several examples are presented to illustrate the performance of both the original and the new methods.

RÉSUMÉ. On considère le problème d'estimer les coefficients d'un système d'équations différentielles quand une trajectoire du système est connue en un petit nombre d'instants. On utilise pour cela une méthode de Monte Carlo très simple, la méthode de rejet qui ne fournit pas directement une estimation ponctuelle des coefficients comme le font les méthodes déterministes mais plutôt un ensemble de valeurs de ces coefficients qui sont cohérentes avec les données. L'examen des propriétés de cette méthode permet de comprendre non seulement comment bien choisir les différents paramètres de la méthode lorsqu'on l'utilise mais aussi d'introduire une méthode plus efficace, en deux étapes, que nous appelons la méthode de rejet séquentielle. Plusieurs exemples illustrent les performances respectives de la méthode d'origine et de la nouvelle méthode.

KEYWORDS : Coefficient fitting, differential system, Monte Carlo method, sequential rejection sampling.

MOTS-CLÉS : Ajustement de coefficients, systèmes différentiels, méthode de Monte Carlo, méthode de rejet séquentielle.

1. Introduction

Many natural phenomena, for example in biology, are modeled using systems of differential equations that usually involve a large number of variables and a large number of unknown coefficients. Often one would like to estimate the coefficients from some known data, for example from the knowledge of one trajectory at several time points. This paper focuses on the estimation of the coefficients of a differential system using a Monte Carlo approach.

There is a large array of available methods to estimate the coefficients of a differential system. One of the most popular class of methods are the *quasi-Newton methods* [4]. It attempts to find a minimum value of an objective function by estimating iteratively the value of the coefficients for which the gradient of this function is the zero vector. Other methods, the *collocation methods* ([6],[8]), estimate the coefficients by first using the known data to construct an approximation of each component of the solution in terms of a basis function expansion. This approximation and the corresponding approximate derivative are then substituted in the differential system, and the coefficients are the ones that minimize the error.

While these methods can be very effective in estimating the coefficients, these methods usually require that the objective function be convex. As such, more recent studies have focused on those methods which require weaker or even no conditions on the objective function. This includes the so-called *stochastic search algorithms* which have become popular in the past decade, particularly the subset of "nature-inspired" algorithms such as genetic algorithms and particle swarm optimization. Our method, while also stochastic in nature, will have a slightly different approach from most of the available methods. Instead of obtaining just one set of "best" coefficients, we shall use a Monte Carlo method to produce a *distribution* of points which are likely to be the true coefficients. This allows us to not only take into consideration the errors and uncertainties in the known data, but at the same time, to provide a point estimate if necessary. We will see that, even if this method is rather naive and really easy to implement, it can already be quite effective.

This paper is organized into five sections. We introduce in Section 2 a simple, yet effective stochastic method known as the *rejection sampling algorithm* to generate a sample of possible coefficients. Then in Section 3, we study some of the properties of the method, and illustrate them using simulations. Section 4 shows how to improve the basic method, and provides an application on a simplified circadian cycle model before the conclusion in Section 5.

2. The rejection sampling algorithm

We will consider the system of l differential equations of the form

$$y' = g(y; \theta) \tag{1}$$

for $y \in \mathbb{R}^l$ with several unknown coefficients $\theta = (\theta^1, \theta^2, ..., \theta^m) \in \mathbb{R}^m$. We assume that we have known values of $y(t)$ at times $T = \{t_0, t_1, ..., t_k\}$. We denote these data as $\bar{y}(T)$, where $\bar{y}(T)$ is defined as follows:

t	t_0	t_1	$\cdots$	t_k
$\bar{y}(t)$	$\bar{y}_0$	$\bar{y}_1$	$\cdots$	$\bar{y}_k$

We shall denote by $y^i(t; \theta)$ (or simply $y^i(t)$), $i = 1, 2, ..., l$ the ith component of $y(t; \theta)$ and similarly $\bar{y}_j^i$, $i = 1, 2, ..., l$ the ith component of $\bar{y}_j$, $j = 0, 1, 2, ..., k$.

The rejection sampling algorithm was introduced in 1984 by Rubin [7] and generalized by Pritchard et al. [5] in 1999 in the context of population genetics. One proceeds as follows: We begin by generating a sample $\{\theta_i^*\}_{i=1,2,...,N}$ of possible values of the coefficients from a prior distribution $\pi_0(\theta)$. We assume that this distribution has a support S_0 which is usually compact. For each element θ^* of the sample, we compute the corresponding solution $y(t; \theta^*)$ of the differential system (1) with initial condition $(t_0, \bar{y}_0)$ and keep the values of $y(T; \theta)$ at times $T = \{t_0, t_1, ..., t_k\}$. Then, for a convenient measure $\rho(\theta^*)$ of the distance between the trajectory corresponding to the value θ^* of the coefficients and the known data $\bar{y}(T)$, we keep θ^* if $\rho(\theta^*) < \epsilon$, where ϵ is a specified threshold constant, and consider it as part of the sample in the posterior; otherwise, we disregard this value. This process can then be repeated N times or until we have kept some chosen number, say n, of accepted values of θ. One can then construct the histogram of a sample of values from this distribution or compute the summary statistics from the sample. The measure

$$\rho(\theta^*) = \rho(\{y(t_i; \theta^*), i = 1, ..., k\}, \{y_i, i = 1, ..., k\})$$

can be, for example, the sum of squares of the Euclidean distance

$$\rho(\theta^*) = \sum_{i=1}^{k} ||y(t_i; \theta^*) - y_i||^2.$$

We shall call the sample produced in this manner a *rejection sample*.

Definition 1. *Let $S_0 \subset \mathbb{R}^m$ and π_0 be a distribution with support S_0. Let $\theta_1, \theta_2, ..., \theta_N$ be an i.i.d. sample from π_0 and let n be the largest ν for which there exists a sequence of integers $(i_k)_{k=1,2,...,\nu}$ such that $1 \leq i_1 < i_2 < ... < i_\nu \leq N$ and $\rho(\theta_{i_k}) < \epsilon$ for all k. We call $\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_n}$ a **rejection sample** of size n. The **acceptance rate** of this sample is given by $\tau_{RS} = n/N$. The **acceptance region** with maximum threshold ϵ is the set*

$$A_\epsilon = \{\theta \in S_0 | \rho(\theta) < \epsilon\}.$$

For simplicity, whenever no confusion arises, we denote a rejection sample of size n as $\theta_1, \theta_2, ..., \theta_n$ instead of $\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_n}$.

REMARK. — Notice that, by construction, for a given ϵ, the rejection sample $\theta_1, \theta_2, ..., \theta_n$ is drawn from the distribution

$$\pi_\epsilon(\theta|\bar{y}) = \frac{\pi_0(\theta)\mathbb{1}_{A_\epsilon}(\theta)}{\int_{A_\epsilon} \pi_0(\theta)d\theta},$$

where A_ϵ is the acceptance region with maximum threshold ϵ. The distribution $\pi_\epsilon(\theta|\bar{y})$ is precisely the posterior distribution for this algorithm.

Example 2. Before studying the properties of rejection samples, we first illustrate how the method works on a very simple differential equation. Suppose we wish to estimate the coefficients r and K in the logistic differential equation

$$y' = ry\left(1 - \frac{y}{K}\right) \tag{2}$$

that best fit the following data:

t	0	1	2	3	4	5	6	7	8	9	10	12	14	18
y	4	7	12	19	28	48	70	103	140	176	205	238	256	265

Even if the coefficients $\theta = (r, K)$ are unknown, assume that we know that $r \in [0, 1]$ and that $K \in [100, 300]$. We then take the uniform distribution over $[0, 1] \times [100, 300]$ as prior distribution π_0 for sampling θ. Assume the value of ϵ is chosen as 1300, which represents an average error of 10 units for each of the 13 time points which are allowed to vary. The method is run for $N = 5000$ iterations, and for this particular run, we ended up with 65 accepted coefficients. The mean of the accepted values is approximately $(0.5349, 267.61)$ while the minimum value of ρ occurred at about $\theta = (0.5351, 265.94)$. In Figure 1(B), we can see that the resulting trajectory after substituting $\theta = (0.5351, 265.94)$ to (2) is very close to the given data. We also notice in Figure 1(A) that, even with a relatively large value of ϵ, the resulting sample is distributed uniformly, as in our prior distribution, in a small elliptical region which is nothing else than A_ϵ.

3. Main properties

In this section, we will first study the size and the shape of the acceptance region, then give different ways to deduce a point estimate of θ from the posterior distribution. Finally, we shall also study the different factors that affect the acceptance rate.

Assume that the observed data are the actual values of a trajectory $y(T; \theta_0)$ where the time points are given by $T = \{t_0, t_1, ..., t_k\}$. In this case, we shall call θ_0 the "true" value of the coefficient. While this may not be consistent with actual data, this assumption allows the study of the accuracy of the method without worrying about the effect of errors in the data.
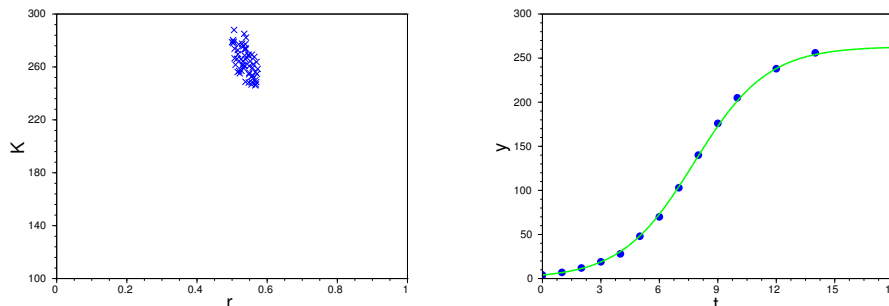
. *(A) Plot of the resulting sample $(r_1, K_1), ..., (r_n, K_n)$ in the rectangle which is the support S_0 of the prior distribution.*

. *(B) Graph of the logistic curve with "minimum distance" coefficients. The known data are indicated by the blue points.*

Figure 1. *Results for logistic model using rejection sampling algorithm*

3.1. The acceptance region

To study the acceptance region, we shall mainly focus on two-variable coefficient estimation problems. Such a simple model will allow us to understand the properties more easily, as this will make the graph of the acceptance region A_ϵ easier to visualize.

The size and shape of the region A_ϵ is clearly dependent on the value of ϵ and the differential system through the function $\rho(\theta)$. Obviously, as soon as ϵ is larger than the maximum of $\rho(\theta)$ for $\theta \in S_0$, the acceptance region will be the entire support S_0. Otherwise, it will be helpful to look at the contour map for $\rho(\theta)$ to understand the shape of the acceptance region better. Consider the two-coefficient Lotka-Volterra model with $\theta = (a, b)$:

$$\begin{cases} \dfrac{dx}{dt} & = & ax - xy \\ \dfrac{dy}{dt} & = & bxy - y \end{cases} \tag{3}$$

Figure 2 gives the contour maps for $\rho(\theta)$ for two different values of θ_0:

One can see that for ϵ small enough, the boundary of A_ϵ looks like an ellipse. It turns out that this is to be expected. As we have assumed in this section that θ_0 is a "true" value of the parameter, by definition $\rho(\theta_0) = 0$ and θ_0 is the minimum of ρ. Moreover, θ_0 belongs to the support of the prior distribution. If we assume that the system is of
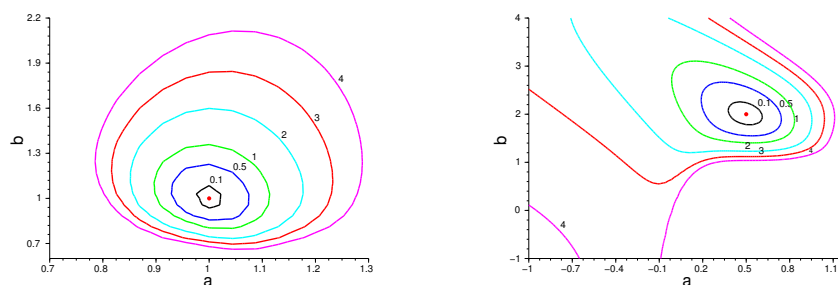
. *(A) True coefficient value:* $\theta_0 = (a, b) = (1, 1)$*, initial point* $(0.5, 1.5)$

. *(B) True coefficient value:* $\theta_0 = (a, b) = (0.5, 2)$*, initial point* $(1, 0.4)$

Figure 2. *Contour maps for the two-coefficient Lotka-Volterra model* (3) *for* $\rho = 0.1, 0.5, 1, 2, 3, 4$. *In both cases, we assume that the known values of* (x, y) *are at the times* $T = \{0, 1, 2, ..., 7\}$, ρ *is the sum of squared differences, and we draw only those values of* (a, b) *within the region* $[-1, 2] \times [-1, 4]$. *The point in red corresponds to* θ_0.

class C^2, then $y(T; \theta)$ and consequently, $\rho(\theta)$, are both of class C^2 as well. Thus, if $\Delta\theta = \theta - \theta_0$ and H is the Hessian of ρ, the ellipse

$$\frac{1}{2}\Delta\theta^T H(\rho(\theta_0))\Delta\theta < \epsilon \tag{4}$$

is an approximation of A_ϵ for ϵ small enough. This implies that the region $\rho < \epsilon$ has the shape of an ellipse around θ_0 as soon as ϵ is small enough.

In contrast, for larger values of ϵ, there is no reason that the acceptance region will be an ellipse, as we can see in Figure 2(B).

REMARK. — Although we did not encounter such a case in our previous example, it is not difficult to see that when ρ has several local minima, the contour plot may consist of multiple closed regions. For examples, one can refer to [1].

3.2. Point estimation

Although the rejection sampling algorithm does not directly provide a point estimate of the unknown coefficients but rather a sample $\theta_1, \theta_2, ..., \theta_n$, a point estimate is easy to compute from this sample. There are at least two natural ways to do this: one can either take the average $\widehat{\theta}_{ave}$ of the entire rejection sample or take $\widehat{\theta}_{min}$, which is the element of the rejection sample that gives the minimum distance with respect to the metric ρ. We shall first explain in this section why, for the case where θ_0 is the true value of the coefficients, the one which produces the minimum distance is usually a better choice.

Then we will introduce a less simple but more efficient estimator that we call the *least mean estimate* which combines an average and a minimum.

Let $\widehat{\theta}_{ave}$ denote the sample average:

$$\widehat{\theta}_{ave} = \frac{1}{n} \sum_{i=1}^{n} \theta_i,$$

where the sum is taken component-wise, and let $\widehat{\theta}_{min}$ denote the minimum distance estimate:

$$\widehat{\theta}_{min} = \operatorname*{ArgMin}_{\theta \in \{\theta_1, \theta_2, \ldots, \theta_n\}} \rho(\theta).$$

By the Law of Large Numbers, we know that $\hat{\theta}_{ave}$ converges almost surely to the expectation of θ with distribution $\pi_0(\theta|\rho(\theta) < \epsilon)$. This need not be equal to θ_0, especially for large ϵ. This is true even for a uniform prior, as the region may not be centered on θ_0. In contrast, $\hat{\theta}_{min}$ has a very nice property as an estimator for θ_0, as we shall now show.

Proposition 3. *Let θ_1, $\theta_2, \ldots, \theta_n$ be a rejection sample for threshold value $\epsilon > 0$ to estimate the coefficients θ in a differential equation $y' = g(y; \theta)$, and let*

$$\hat{\rho}_n(\theta) = min\{\rho(\theta_1), \rho(\theta_2), \ldots, \rho(\theta_n)\}.$$

Assume also that the prior distribution π_0 for the rejection sample is absolutely continuous. If $\theta_0 = \operatorname*{ArgMin}_{\theta \in S_0} \rho(\theta)$, then $\hat{\rho}_n(\theta) \xrightarrow{\mathcal{P}} \rho(\theta_0)$ as the sample size n tends to ∞.

Proof. Let ρ_i denote the distance $\rho(\theta_i)$. For any $K > 0$, first note that, as $\theta_1, \ldots, \theta_n$ is a rejection sample and thus i.i.d., we have

$$\mathbb{P}(\hat{\rho}_n(\theta) \leq K) = 1 - [\mathbb{P}(\rho_1 \geq K)]^n. \tag{5}$$

Indeed:

$$
\begin{aligned}
\mathbb{P}(min(\rho_1, \rho_2, \ldots, \rho_n) \leq K) &= 1 - \mathbb{P}(min(\rho_1, \rho_2, \ldots, \rho_n) \geq K) \\
&= 1 - \mathbb{P}(\rho_1 \geq K, \rho_2 \geq K, \ldots, \rho_n \geq K) \\
&= 1 - [\mathbb{P}(\rho_1 \geq K)]^n
\end{aligned}
$$

Thus, for any $\alpha > 0, n > 0$,

$$
\begin{aligned}
\{|\hat{\rho}_n(\theta) - \rho(\theta_0)| < \alpha\} &= \{\rho(\theta) - \alpha < \hat{\rho}_n(\theta) < \rho(\theta_0) + \alpha\} \\
&= \{\hat{\rho}_n < \rho(\theta_0) + \alpha\} \setminus \{\hat{\rho}_n \leq \rho(\theta_0) - \alpha\}
\end{aligned}
$$

Since $\{\hat{\rho}_n < \rho(\theta_0) - \alpha\} \subset \{\hat{\rho}_n \leq \rho(\theta_0) + \alpha\}$, we have

$$
\begin{aligned}
\mathbb{P}(|\hat{\rho}_n(\theta) - \rho(\theta_0)| < \alpha) &= \mathbb{P}(\hat{\rho}_n < \rho(\theta_0) + \alpha) - \mathbb{P}(\hat{\rho}_n \leq \rho(\theta_0) - \alpha) \\
&= [\mathbb{P}(\rho_1 > \rho(\theta_0) - \alpha)]^n - [\mathbb{P}(\rho_1 \geq \rho(\theta_0) + \alpha)]^n, \quad (6)
\end{aligned}
$$

where the last line follows from (5). The limit of the first term of (6) as $n \to \infty$ is 1 since by definition, $\rho(\theta_0)$ is the minimum. On the other hand, the limit of the second term is 0 since $\mathbb{P}(\rho_1 \geq \rho(\theta_0) + \alpha) < 1$ from the absolute continuity of $\pi_0(\theta)$. Thus, $\lim_{n \to \infty} \mathbb{P}(|\hat{\rho}_n(\theta) - \rho(\theta_0)| < \alpha) = 1$, and so $\hat{\rho}_n(\theta)$ converges in probability to $\rho(\theta_0)$. $\qquad \square$

As an illustration of Proposition 3, let us consider a simple example taken from [1] that shows that increasing the number of iterations N does not improve $\widehat{\theta}_{ave}$ but significantly improves $\widehat{\theta}_{min}$. In Tables 1 and 2, we show how the resulting estimates $\widehat{\theta}_{ave}$ and $\widehat{\theta}_{min}$ vary when estimating the coefficients $(\alpha, \alpha_0, \gamma, \beta)$ in the repressilator model of Elowitz and Liebler [3] given by

$$
\begin{cases}
\dfrac{dm_1}{dt} &= -m_1 + \dfrac{\alpha}{1 + p_3^\gamma} + \alpha_0 \\[2mm]
\dfrac{dp_1}{dt} &= -\beta(p_1 - m_1) \\[2mm]
\dfrac{dm_2}{dt} &= -m_2 + \dfrac{\alpha}{1 + p_1^\gamma} + \alpha_0 \\[2mm]
\dfrac{dp_2}{dt} &= -\beta(p_2 - m_2) \\[2mm]
\dfrac{dm_3}{dt} &= -m_3 + \dfrac{\alpha}{1 + p_2^\gamma} + \alpha_0 \\[2mm]
\dfrac{dp_3}{dt} &= -\beta(p_3 - m_3)
\end{cases}
\qquad (7)
$$

We increase the number of iterations N while holding everything else constant. Assume that the true values of the coefficients are $(\alpha, \alpha_0, \gamma, \beta) = (1000, 1, 2, 5)$ and $\epsilon = 2000$. Looking at the results in Table 1, we can see that $\hat{\rho}_N(\hat{\theta}_{ave})$ does not decrease for the estimates using $\widehat{\theta}_{ave}$ as N increases. However, the values of $\hat{\alpha}_{ave}, \hat{\alpha}_{0 ave}, \hat{\gamma}_{ave}$ and $\hat{\beta}_{ave}$ are consistently around 1010, 1.30, 2.05, and 5.8, respectively, which we can assume to be close to the actual expectation of the distribution $\pi(\theta | \rho(\theta) < 2000)$. In contrast, the results in Table 2 show that while the values of $\hat{\rho}_N(\hat{\theta}_{min})$ may not always decrease monotonically as N increases (due to the random nature of the sample), there is still a clear trend of decrease for ρ as $N \to \infty$.

Even if the choice of the minimum distance estimate is usually better than the average, we lose in taking the minimum distance estimate the benefit of taking an average between several good estimates in order to average the different errors and get something in the interior of all the estimates. This observation gives the idea of a new way to deduce an estimate of the coefficient from the rejection sample, that we will explain now and illustrate with an example.

Table 1. *Results of* $\widehat{\theta}_{ave} = (\hat{\alpha}_{ave}, \hat{\alpha}_{0ave}, \hat{\gamma}_{ave}, \hat{\beta}_{ave})$ *for* $\epsilon = 2000$. *True value:* $\alpha = 1000, \alpha_0 = 1, \gamma = 2, \beta = 5$*, prior distribution: uniform over* $[800, 1200] \times [0, 4] \times [0, 7] \times [0, 10]$

N	α	α_0	γ	β	$\hat{\rho}_N(\hat{\theta}_{ave})$
500	1004.3	1.1800	2.0730	6.1476	265.03
1000	1008.0	1.3508	2.0534	5.6957	337.78
2000	1005.1	1.3364	2.0394	5.7085	412.41
4000	1022.3	1.2380	2.0354	5.8069	342.04
8000	1010.5	1.3224	2.0521	5.8860	403.34
16000	1011.0	1.2886	2.0409	5.8135	390.4

Table 2. *Results of* $\widehat{\theta}_{min} = (\hat{\alpha}_{min}, \hat{\alpha}_{0min}, \hat{\gamma}_{min}, \hat{\beta}_{min})$ *for* $\epsilon = 2000$. *True value:* $\alpha = 1000, \alpha_0 = 1, \gamma = 2, \beta = 5$*, prior distribution: uniform over* $[800, 1200] \times [0, 4] \times [0, 7] \times [0, 10]$

N	$\hat{\alpha}_{min}$	$\hat{\alpha}_{0min}$	$\hat{\gamma}_{min}$	$\hat{\beta}_{min}$	$\hat{\rho}_N(\hat{\theta}_{min})$
500	1158.6	0.2709	1.7097	3.8584	352.06
1000	1142.2	0.7723	1.9288	4.8699	138.67
2000	914.4	0.2616	1.801	4.0841	136.64
4000	819.77	0.9172	2.0406	5.143	37.59
8000	845.32	0.9282	2.0363	5.0312	38.71
16000	859.59	0.9781	2.0452	5.1839	26.57

The main idea is the following: if we average the d best values of the coefficients θ in terms of the distance ρ, and if we increase the number of values in this average, the result will initially be located around θ_0 before it tends, as d increases, towards the mean of the distribution $\pi(\theta | \rho(\theta) < \epsilon)$. In fact, taking the mean of these few initial θ's can often give a better result than just choosing the one with the minimum distance.

More precisely, let $\theta_1, \theta_2, ..., \theta_n$ be a rejection sample and denote by $\rho(\theta_{(1)}), \rho(\theta_{(2)}), ..., \rho(\theta_{(n)})$ the distances $\rho(\theta_1), \rho(\theta_2), ..., \rho(\theta_n)$ arranged in increasing order. This consequently defines the reordering of the sample from the best (lowest ρ) to the least $\theta_{(1)}, \theta_{(2)}, ..., \theta_{(n)}$. Let

$$\bar{\theta}_d = \frac{1}{d} \sum_{i=1}^{d} \theta_{(i)}$$

for some d, $1 \leq d \leq n$. The *least mean estimator* is given by

$$\widehat{\theta}_{leastmean}(d) = \underset{\theta \in \{\bar{\theta}_1, ..., \bar{\theta}_d\}}{\text{ArgMin}} \rho(\theta). \tag{8}$$

After computing $\widehat{\theta}_{leastmean}(d)$ for a large number of numerical experiments, it has been observed that in most cases, this estimator produces coefficients with the least distance when $d = 10$. Note that $\hat{\theta}_{min}$ is equal to $\widehat{\theta}_{leastmean}(1)$.

Example 4. We apply this choice of best estimate to compute a best guess of θ in the simplified Lotka-Volterra model introduced in (3). Figure 3 shows the evolution of $\widehat{\theta}_{leastmean}(d)$ in estimating the coefficients of a Lotka-Volterra model as d increases.
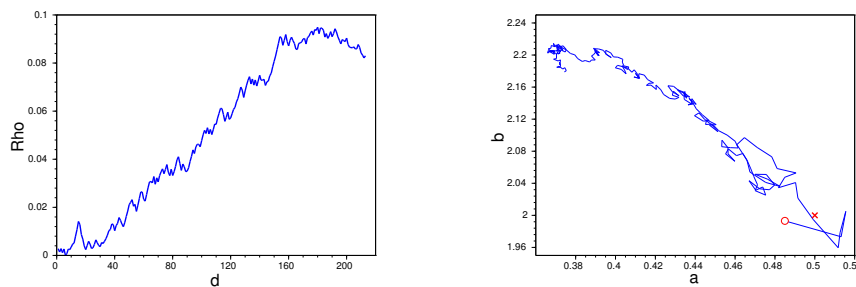
. (A) Typical graph of $\rho(\widehat{\theta}_{leastmean}(d))$ as d increases

. (B) Evolution of $\widehat{\theta}_{leastmean}(d)$ as d increases. The red circle represents the minimum distance point, and the red 'x' are the model coefficients.

Figure 3. Evolution of $\widehat{\theta}_{leastmean}(d)$ as the number of elements d increases for the coefficients in a Lotka-Volterra model with model coefficients $(a, b) = (0.5, 2)$. The initial point of the model data is at $(1, 0.5)$ and the time points are at $T = \{0, 0.5, 1, ..., 3\}$. The prior distribution is uniform over $[0, 3] \times [0, 3]$.

In Figure 3(A), we observe that $\rho(\widehat{\theta}_{leastmean}(d))$ initially decreases, and then increases until it stabilizes at a certain point. By the Law of Large Numbers, this point is the value of ρ for the center of A_ϵ. Figure 3(B) illustrates why $\rho(\widehat{\theta}_{leastmean}(d))$ decreases initially before it increases again. The coefficients $\overline{\theta}_d$ are initially located around θ_0 for small values of d, before they tend towards the center of A_ϵ as d increases.

3.3. The acceptance rate

As we have seen in the previous section, the larger the resulting rejection sample, the higher the chances of obtaining a more accurate estimate for θ. In the next three subsections, we will see how the acceptance rate is affected by the sample size N, the maximum threshold ϵ, and the number of coefficients m.

3.3.1. Dependence on the sample size

Let $\theta_1, \theta_2, ..., \theta_N$ be an i.i.d. sample from π_0. Define the corresponding random variables $I_i = \mathbb{1}_{\{\theta_i \in A_\epsilon\}}, i = 1, 2, ..., N$. Then, $I_1, I_2, ..., I_N$ are i.i.d. Bernoulli random variables with success probability

$$\tau = \mathbb{P}(\theta \in A_\epsilon) = \int_{A_\epsilon} \pi_0(\theta)d\theta. \tag{9}$$

Thus, by the Law of Large Numbers, as $N \to \infty$, the rejection sampler acceptance rate $\tau_{RS} = \frac{1}{N}\sum_{i=1}^{N} I_i$ converges in probability to τ.

REMARK. — When the prior distribution π_0 is uniform, (9) implies that the acceptance rate is simply the ratio of the volumes of A_ϵ and S_0.

Furthermore, by the Central Limit Theorem,

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} I_i - \tau\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau(1-\tau)).$$

This means that provided N is large enough, a 95% confidence interval for the acceptance rate τ_{RS} is approximately given by $\tau \pm 1.64\sqrt{\frac{\tau(1-\tau)}{N}}$. Since the maximum of $\tau(1-\tau)$ occurs when $\tau = 0.5$, a conservative estimate for the length of the confidence interval is $1.64/\sqrt{N}$. In general, τ cannot be computed directly, and we will use its estimator τ_{RS} instead. For large N, this confidence interval implies that we can actually expect the acceptance rate to be more or less the same in each sample, and to be on a narrow range around τ.

3.3.2. Dependence on ϵ

Although the sample size N may not play a large role on the acceptance rate of our sample, it is clear that this is not the same with the maximum threshold ϵ. In fact, for any sample $\theta_1, \theta_2, ..., \theta_N$, the asymptotic acceptance rate τ is a nondecreasing function of ϵ. To see why this is so, denote by τ_ϵ^N the acceptance rate associated with the threshold ϵ. If $\epsilon_1 < \epsilon_2$, then $A_{\epsilon_1} \subset A_{\epsilon_2}$. Thus, $\tau_{\epsilon_1}^N = \mathbb{P}(\theta \in A_{\epsilon_1}) \le \mathbb{P}(\theta \in A_{\epsilon_2}) = \tau_{\epsilon_2}^N$. This inequality also holds for the asymptotic acceptance rate.

When the prior distribution is uniform, we can say even more about τ. In particular, it can be shown that the rate of increase of the acceptance rate is of the order $l/2$, where l is the dimension of the coefficient space.

Proposition 5. *Suppose that the rejection method is used to estimate θ in the differential system $y' = g(y; \theta)$, where the known data $\bar{y}$ consists of the actual points for $y(t; \theta)$ when*

$\theta = \theta_0$. *Assume also that θ is l-dimensional, and that the prior distribution is uniform on its support. Then, the acceptance rate τ_{RS} satisfies*

$$\lim_{N \to \infty} \tau_{RS} = \frac{(2\pi\epsilon)^{l/2}}{\Gamma(\frac{l}{2} + 1)\sqrt{\prod_{i=1}^{l} \lambda_i}}, \tag{10}$$

where $\lambda_1, \lambda_2, ..., \lambda_l$ are the eigenvalues of the Hessian matrix of $\rho(\theta_0)$ and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du.$$

For a proof, one can refer to Proposition 3.12 in [1].

Example 6. To illustrate this proposition, we return to the logistic model in Example 2 as it simple enough to allow explicit computation. Recall that the best fit coefficients were approximately $r_0 = 0.5351$ and $K_0 = 265.94$ and the prior distribution was chosen to be uniform on the rectangle $[0, 1] \times [100, 300]$. The objective function can be computed exactly as

$$\rho(r, K) = \sum_{i=0}^{13} \left(\frac{4Ke^{rt_i}}{K - 4 + 4e^{rt_i}} - \bar{y}_i \right),$$

where (t_i, y_i) are the 14 data points which were previously given. Assuming that $\theta_0 = (r_0, K_0)$, we shall use $\epsilon - \rho(\theta_0) \approx 1266.2$ on the right-hand side of (4). Using a symbolic computation program such as Maple, one can compute the Hessian of ρ at θ_0:

$$H = \begin{pmatrix} 2516643.593 & 2220.7277 \\ 2220.7277 & 6.264013 \end{pmatrix}, \tag{11}$$

which has eigenvalues $\lambda_1 = 4.3044$ and $\lambda_2 = 2516645.6$. This means that the semi-principal axes are $\sqrt{2532.4/4.3044} \approx 24.255$ and $\sqrt{2532.4/25616645.6} \approx 0.03172$. If we assume that the prior distribution π_0 is uniformly distributed on its support, the acceptance rate is thus $\tau = \pi(24.255)(0.03172)/200 \approx 0.01208$. Since the acceptance rate that we obtained from the sample in Example 2 was $65/5000 = 1.3\%$ and the number of iterations was $N = 5000$, the 95% confidence interval for τ is thus

$$0.013 \pm 1.645 \sqrt{\frac{(0.013)(0.987)}{5000}} = (0.01036, 0.01564),$$

which contains the asymptotic acceptance rate $\tau = 0.01208$.

Another illustration of the previous proposition is given by Table 3 below. The table shows the average acceptance rate for five runs when estimating the coefficients of the repressilator model (see (7)). For the same data, ϵ, and prior distribution, and for a sample

size of $N = 5000$, we can see, as expected, that as ϵ decreases, τ also decreases. Furthermore, we also observe that when ϵ is halved from 1200 to 600, the acceptance rate is reduced to about one-fourth, from 8.38% to 2.33%, which is on line with Proposition 5 which suggests that the rate of decrease in the acceptance rate is of order 2.

Table 3. *Average acceptance rate for the repressilator model as ϵ varies*

ϵ	Average acceptance rate
1200	8.38%
1000	5.81%
800	3.72%
600	2.33%

3.3.3. Dependence on the dimension of the coefficient space

Aside from the size of ϵ, the dimension of the coefficient space also plays an important role in the acceptance rate. As Proposition 5 implies, it is clear that if θ is of large dimension, one will get a low acceptance rate. We illustrate this in the following example.

Example 7. Consider again the repressilator model for gene regulatory networks (see (7)). We assume that the initial point is at $(0, 2, 0, 1, 0, 3)$ and that there is a "true" value of the coefficients, namely $(\alpha_0, \gamma, \beta, \alpha) = (1, 2, 5, 1000)$. Seven points were chosen to represent the trajectory, in particular at $t = 0, 1, 2, 3, 4, 5, 6$. The prior distribution is uniformly distributed on $[0, 2] \times [1.5, 2.5] \times [0, 10] \times [900, 1100]$. We produced 10000 iterations, and kept only those which yield trajectories having a maximum distance of 60. The number of unknown coefficients is different in each experiment. As the acceptance rate varies on each experiment, we take the average of the acceptance rate when producing five separate rejection samples. The results are given in Table 4, together with the corresponding acceptance rate.

Table 4. *Acceptance rate of the repressilator when the number of unknown coefficients varies. The results shown are the average acceptance rate for five separate runs.*

Coefficients fixed	Average acceptance rate (in percent)
None	0.404
α	0.474
α, α_0	1.372
α, α_0, γ	10.2

We can see from Table 4 that as the dimension of the coefficient space increases, the corresponding acceptance rate decreases, where it is less than 1 percent for the case of 3

or more variables. This is even if the support of the prior distribution that we have chosen covers a region which is a small neighborhood of θ_0. This means that one will need a larger sample to obtain a relatively accurate estimate of θ. We will see in the next section one way to alleviate this situation.

4. Sequential rejection method

Although the method which we have described is quite simple, one can quickly realize that there are a large number of parameters that need to be chosen when implementing it. Without a careful choice of these parameters, one can easily obtain a rejection sample of size 0. Among the main choices, the choices of the prior distribution and of the value of ϵ are the most difficult. In all the previous examples, we have chosen the prior to be a uniform distribution on a cartesian product of intervals, which is realistic, for example, when the user of the model knows a range of values for each of the coefficients. But if the support of this distribution is too narrow, there is a big risk that it will not contain any "good" values. On the other hand, if it is too large, sampling will be like finding a needle in a hay stack. The same kind of difficulty exists for the choice of ϵ.

In this section, we propose an improved version of the rejection sampling algorithm we call the *sequential rejection method* which mainly consists in choosing an improved prior distribution based on the result of a preliminary sample and an adequate choice of ϵ based on the sensitivity of the solution of the differential system with respect to the different unknown coefficients. We will explain how one can choose the prior distribution and the threshold ϵ, provide a summary of the new algorithm, and then finally apply it on an example.

4.1. Choice of the prior

A good way to choose the prior properly is to do it in two steps. First, begin with a uniform prior for, say, the first 10% of the desired number of iterations. Provided this gives a sufficiently large number of accepted values so that it can adequately represent the acceptance region A_ϵ, we can compute the mean vector μ and the covariance matrix Σ for this sample. Then, for the remaining 90% of the sample, one can change the initial prior to a multivariate Gaussian distribution with mean and covariance matrix equal to that of the preliminary sample.

As we shall see in the next example, using this two-step approach substantially increases the acceptance rate and provides the same acceptance rate even when ϵ is decreased. This therefore gives a much higher accuracy of the resulting sample.

Example 8. Consider the repressilator model (see (7)) with "true" coefficients $(\alpha, \alpha_0, \gamma, \beta) = (1000, 1, 2, 5)$, initial point $(0, 2, 0, 1, 0, 3)$, $T = \{0, 0.5, 1, ..., 3\}$ and $N = 5000$ iterations. One can observe from the results in Table 5 that the average among the five runs of the minimum ρ decreases as ϵ decreases. This is not completely surprising, as we are

obtaining approximately the same number of points in a region which contains less points of high distance ρ. Thus, we would expect the minimum distance to also decrease.

Table 5. *Results using sequential rejection, $N = 5000$, where the first 500 are used as a preliminary sample. The prior distribution is uniform over $[0, 7] \times [0, 4] \times [0, 10] \times [800, 1200]$, and the average acceptance rate of 5 runs for the last 4500 iterations are shown below.*

ϵ	Min. distance	Acceptance rate
1200	8.97	66.04
1000	7.67	65.23
800	5.26	62.67
600	4.6	62.45

One interesting observation is that the acceptance rate for the second part of the sample seems to remain approximately the same in the sequential method, regardless of the value of ϵ. It turns out that this is true in general, as we shall show in Corollary 10. For this, we first need the following technical result.

Proposition 9. *Let Y be a random vector of dimension m whose components are the coordinates of a point chosen uniformly within and on the ellipse $y^T A y < 2\epsilon$, where A is positive definite and ϵ is constant. Then the covariance matrix Σ_Y of Y is equal to $\frac{2\epsilon}{m+2} A^{-1}$.*

Proof. As the ellipse $y^T A y < 2\epsilon$ is a linear transformation of the unit ball $x^T x < 1$ by $y = \sqrt{2\epsilon} A^{-1/2} x$, it is easy to verify that the proof of the proposition reduces to that of showing that for a random vector $X = (X_1, X_2, ..., X_m)$ whose components are the coordinates of a point chosen uniformly within the m-dimensional unit ball $x^T x < 1$, the covariance matrix of X is given by $\Sigma_X = \frac{1}{m+2} I_m$, where I_m is the $m \times m$ identity matrix. As the support is symmetric about the origin, the mean of each coordinate and the covariance between any two coordinates must be 0. By the exchangability of the X_i's and integrating in hyperspherical coordinates, it can be shown that the variance of each X_i must be $\frac{1}{m+2}$. Combining these facts, we can thus conclude that the covariance matrix of X is $\Sigma_X = \frac{1}{m+2} I_m$. $\square$

An interesting corollary of the previous technical result is that, under some conditions, the acceptance rate for the second part of the sample as we described above, is actually constant even if ϵ is decreased, as we shall now show.

Corollary 10. *Suppose that the acceptance region A_ϵ is exactly a hyperellipsoid, and that the uniform distribution on A_ϵ has covariance matrix Σ. If Y is drawn from another prior distribution with covariance matrix Σ, then $\mathbb{P}(Y \in A_\epsilon)$ is independent of ϵ.*

Proof. First note that μ and Σ are the mean vector and covariance matrix of a uniform distribution within the hyperellipsoid $\frac{1}{2}(\Delta\theta)^T H(\rho(\theta))\Delta\theta < 2\epsilon$ defined in (4). This means that by Proposition 9, $\Sigma = \frac{2\epsilon}{m+2}H^{-1}$, or that $H = \frac{2\epsilon}{m+2}\Sigma^{-1}$. Thus, the probability of an arbitrary Y to be within the acceptance region is given by

$$
\begin{aligned}
\mathbb{P}[(\Delta\theta)^T H\Delta\theta < 2\epsilon] &= \mathbb{P}\left[(\Delta\theta)^T \frac{2\epsilon}{m+2}\Sigma^{-1}\Delta\theta < 2\epsilon\right] \\
&= \mathbb{P}[(\Delta\theta)^T\Sigma^{-1}\Delta\theta < m+2],
\end{aligned}
$$

which is independent of ϵ. $\qquad\square$

If we assume that the sample covariance is not too far from Σ and that A_ϵ is close to a hyperellipsoid, then Corollary 10 implies that the strategy of sampling with any distribution with covariance matrix Σ will give the same acceptance rate regardless of the value of ϵ. Thus, if the new distribution is chosen well, this can allow us to obtain a high acceptance rate even if the value of ϵ is small. One possible choice is the multivariate Gaussian distribution.

While the method provides a significant improvement from the acceptance rate of the basic rejection sampling algorithm, it also suffers from some problems. In practice, we do not know what the shape of the acceptance region A_ϵ is. If A_ϵ is too far from an ellipsoid, the Gaussian prior in the second part of the sample may not be very effective in increasing the acceptance rate. Secondly, this method is dependent on having a good estimate of Σ. However, if the size of the initial sample is not large enough, the covariance matrix S of this sample may not approximate Σ very well. Furthermore, increasing the size of the sample is not a guarantee of an improved covariance matrix estimate as S is known as an inconsistent estimator for Σ. Despite these problems, performing this sequential rejection method is still a much better alternative than using the basic rejection method.

4.2. Choosing the value of ϵ

Another important consideration when using the method is choosing an appropriate value of the maximum threshold ϵ. We now provide a logical way to choose ϵ based on the support of our prior distribution and the differential system which we shall use to model the data.

To obtain a reasonable value of ϵ, one needs to have an idea of the order of magnitude of $\rho(\theta)$. Suppose that, as before, we take as measure of distance the sum of squared differences

$$
\rho(\theta) = \sum_{j=0}^{k}(y(t_j;\theta) - \bar{y}_j)^2,
$$

where $\bar{y}$ is assumed to be the model data for $\theta = \theta_0$. Our objective is to construct an estimate of a "typical value" of $\rho(\theta)$.

The main idea of our approximation for ϵ involves computing an estimate of

$$y(t_j; \theta) - \bar{y}_j = y(t_j; \theta) - y(t_j; \theta_0)$$

for a fixed time t_j. Since θ is m-dimensional, a simple way to do this is to use the total differential

$$y(t_j; \theta) - y(t_j; \theta_0) \approx \sum_{i=1}^{m} \left\{ \frac{\partial y}{\partial \theta^i}(\theta_0) \right\} \Delta\theta^i, \tag{12}$$

where the superscript i denotes the ith component of the coefficient vector and $\Delta\theta^i = (\theta^i - \theta_0^i)$. The partial derivative in (12) can be estimated using a finite difference estimate

$$\epsilon_{ij} = \frac{y(t_j; \theta_0^1, \theta_0^2, ..., \theta_0^i + h, ..., \theta_0^m) - y(t_j; \theta_0^1, ..., \theta_0^m)}{h}, \tag{13}$$

where h is a small increment (for example, around 10^{-6}).

We are then left with the choice of θ_0 and $\Delta\theta^i = \theta^i - \theta_0^i$ to be used in the computation. In general, we have no idea of what θ_0 is exactly. The most logical guess we could make is that it is at the center of the support of the prior distribution. Thus, assuming that our prior distribution is a box $(\theta_1^{min}, \theta_1^{max}) \times (\theta_2^{min}, \theta_2^{max}) \times ... \times (\theta_m^{min}, \theta_m^{max})$ on $\mathbb{R}^m$, this gives us the estimate

$$\theta_0^* = \left(\frac{\theta_1^{min} + \theta_1^{max}}{2}, ..., \frac{\theta_m^{min} + \theta_m^{max}}{2} \right).$$

On the other hand, we can think of the size of $\Delta\theta^i$ as an estimate of the maximum difference we are willing to accept between the ith component of θ and the true θ_0. This can be chosen as a fraction p of the length of the interval for that coefficient. That is,

$$\Delta\theta^{i,*} = p(\theta_{max}^i - \theta_{min}^i).$$

Based on our experiments, we recommend choosing p to be between 0.1 and 0.2.

Using the above choices, we are able to obtain an estimate for $y(t_j; \theta) - \bar{y}_j$. Substituting this estimate in $\rho(\theta)$, we end up with the following estimate for ϵ:

$$\epsilon_{est} = \left(\sum_{j=0}^{k} \sum_{i=0}^{m} \epsilon_{ij} \Delta\theta^{i,*} \right)^2 \tag{14}$$

Example 11 illustrates how ϵ_{est} can be computed in the case of the repressilator (7).

Example 11. Consider the repressilator model (7), where we assume that the initial point is at $(0, 2, 0, 1, 0, 3)$ and the model coefficients to be $(\alpha, \alpha_0, \gamma, \beta, \alpha) = (1000, 1, 2, 5)$. Suppose the known data is at the times $T = \{0, 1, 2, ..., 7\}$, and that the prior distribution is uniformly distributed on $[0, 2] \times [1.5, 2.5] \times [0, 10] \times [900, 1100]$. Also, we choose

$p = 0.1$ Then $\theta_0^* = (1, 2, 5, 1000)$ and the value of $\Delta\theta^{i,*}$ for $i = 1, 2, 3, 4$ are 0.2,0.1,1, and 20, respectively. A series of finite difference computations would then allow us to compute $\epsilon_{est} \approx 865$. This gives approximately a 1% acceptance rate.

4.3. Implementing the method

We now provide a short summary of the steps needed to apply the sequential rejection method to produce a distribution of coefficients of a differential system that best fits some known data.

1) Choose a preliminary prior distribution π_0 for the unknown coefficients and an initial sample size.

2) Choose a proper ϵ using the procedure outlined in Section 4.2, and choose the percentage of the range to be around 10 to 20 percent. The resulting acceptance rate will vary depending on the differential system, but empirical results suggest it will be at least 3%.

3) Provided we obtain a reasonably large rejection sample in the previous step, change the prior distribution to a Gaussian distribution, with a mean equal to the sample mean and covariance matrix equal to a fraction of the sample covariance (around 10% would be a typical choice).

4) After we obtain the rejection sample $\theta_1, \theta_2, ..., \theta_n$, the scattermatrix of these samples can be obtained to visualize the estimated law of θ. However, if a point estimate is required, arrange the θ_i's in increasing order based on the metric ρ. Then compute $\widehat{\theta}_{leastmean}(10)$.

4.4. Application of the method

We shall now use the rejection sampling method and the steps outlined in Section 4.3 to estimate the coefficients in a more complicated example. Here we estimate the coefficients of the simplified circadian cycle model

$$\begin{cases} \dfrac{dp_1}{dt} &= v_1 \dfrac{K^\gamma}{K^\gamma + c_2^\gamma} - k_3 p_1 p_2 + k_4 c_1 - kd_1 p_1 \\ \dfrac{dp_2}{dt} &= v_2 \dfrac{K^\gamma}{K^\gamma + c_2^\gamma} - k_3 p_1 p_2 + k_4 c_1 - kd_2 p_2 \\ \dfrac{dc_1}{dt} &= k_3 p_1 p_2 - k_4 c_1 - k_1 c_1 + k_2 c_2 - kd_3 c_1 \\ \dfrac{dc_2}{dt} &= k_1 c_1 - k_2 c_2 - kd_4 c_2 \end{cases} \tag{15}$$

introduced by Comet et al. in [2]. This system consists of 4 differential equations and 12 variables $(K, \gamma, k_1, k_2, k_3, k_4, kd_1, kd_2, kd_3, kd_4, v_1, v_2)$.

We begin by generating the model data for the differential system (15). The coefficients chosen were $K = 0.4, \gamma = 15, k_1 = 0.08, k_2 = 0.06, k_3 = 0.08, k_4 =$

$0.06, kd_1 = 0.05, kd_2 = 0.05, kd_3 = 0.05, kd_4 = 0.45, v_1 = 2, v_2 = 2.2$. Here, we obtain oscillations which have decreasing amplitudes, as the value of kd_4 is greater than 0.412182, the bifurcation point as computed in [2]. However, the results remain largely unchanged even if we choose $kd_4 < 0.412182$. Graphing the solution of the differential equation for the chosen coefficients, one can show that the first complete oscillation occurs after around every 12 units. Since there is no reason to believe that the data which we will obtain will correspond to exactly one "period", we assume that we are given the discrete trajectory at time points $0, 1, 2, ..., 7$, which correspond to around half a period. We incorporate a small measurement error by adding a Gaussian term with mean 0 and standard deviation 0.15 for the first three variables, and mean 0 and standard deviation 0.05 for the last variable. The standard deviations were chosen to be different for each variable to take into account the approximate sizes of the values of each variable. Finally, the prior distribution is chosen to be uniform, where the lower and upper bounds are given in the table below:

Variable	Minimum in prior	Maximum in prior	Chosen coefficients
K	0	1	0.4
γ	0	20	15
k_1	0	0.2	0.08
k_2	0	0.2	0.06
k_3	0	0.2	0.08
k_4	0	0.2	0.06
kd_1	0	0.2	0.05
kd_2	0	0.2	0.05
kd_3	0	0.2	0.05
kd_4	0	0.5	0.45
v_1	0	5	2
v_2	0	5	2.2

We shall use the guidelines we proposed in Section 4.3 in running the rejection and sequential rejection method. First, we shall choose $\epsilon \approx 27$, which corresponds to the value obtained when we use the total differential estimate in Section 4.2 with θ as the center of the support, $h = 10^{-6}$, and $p = 0.12$. We shall do two sets of ten runs, the first set being a basic rejection method with 10000 iterations. For the second set of ten runs, we shall use the sequential rejection method introduced in the previous section, with an initial sample of size 2000. We then run another 10000 iterations, but now with a multivariate Gaussian distribution with mean and covariance equal to the corresponding mean and covariance of the initial sample. We shall take as point estimate $\widehat{\theta}_{leastmean}(10)$, where we chose the minimum among the first 10 averages.

The results using both the basic rejection method and the sequential rejection method are given in Table 6. Due to the the substantial improvement in the acceptance rate in the sequential rejection method, we are able to obtain an estimate for the unknown coefficients with a smaller distance with respect to ρ.

Table 6. *Results when applying the rejection method, with and without the sequential improvement, to estimate the coefficients in Comet's simplified circadian cycle model. For each of the ten runs, $\widehat{\theta}_{leastmean}(10)$ and the corresponding distance ρ is computed.*
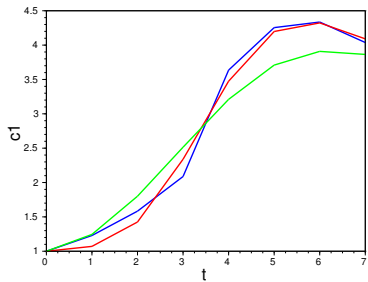
	Sequential rejection for $\epsilon = 27$	Basic rejection for $\epsilon = 27$
Average acceptance	50.46%	3.487%
Average distance	2.13	3.25
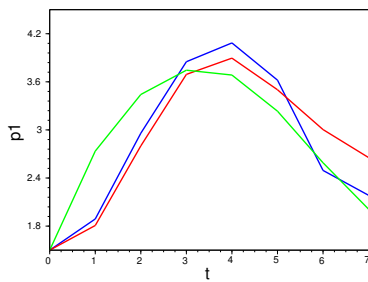
. *(A) Graph of t vs. p_1*
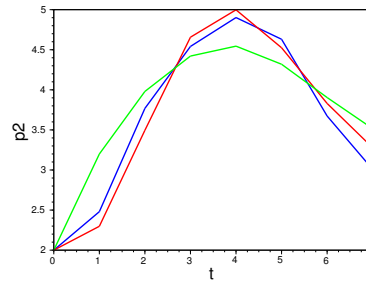
. *(B) Graph of t vs. p_2*

. *(C) Graph of t vs. c_1*
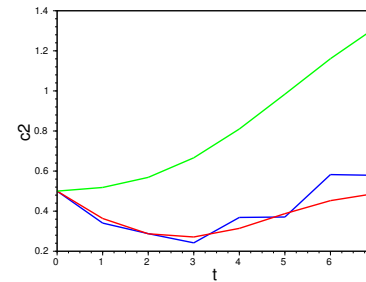
. *(D) Graph of t vs. c_2*

Figure 4. *Graphs of the resulting trajectories using the worst and best coefficients (as given in Table 6) using the sequential rejection method. The graphs in red are using the "best" coefficients, the ones in green are the "worst" coefficients, and the ones in blue are those of the perturbed data.*

An even more interesting comparison is provided by Figure 4, where we graph the trajectories resulting from the "best guess" coefficient estimates using the sequential rejection method with the actual data. We graph the trajectories resulting from two differ-

ent values of θ obtained from the 10 runs: the value of $\widehat{\theta}_{leastmean}(10)$ that produced the smallest ρ (in red), and the one which produced the biggest ρ (in green). These represent the best and worst scenarios when estimating the coefficients using the sequential rejection method. We can see that, even with the limited amount of information provided by eight points of data, the method can give coefficients that fit the known data well, even in a twelve dimensional problem.

5. Conclusion

In this paper, we have studied a Monte Carlo method to compute a distribution of the coefficients of a system of differential equations that fits a known trajectory. Such an approach allows us to incorporate variability in the model as a result of either measurement error, or the variability in individual characteristics in the subjects being studied. Assuming that a true value of the coefficients exists, we have seen that for low dimension problems and a small enough ϵ, we obtain a posterior distribution which allows us to compute good point estimates of θ. However, this method quickly becomes inefficient when the dimension of the coefficient space increases (the "curse of dimensionality") or the chosen prior is too different from the acceptance region because of the low acceptance rate. We have also introduced a new version of this Monte Carlo method, sequential rejection sampling, and explained why it produces a much higher acceptance rate and more accurate estimates.

Note that other sampling methods, like Markov chain Monte Carlo or Sequential Monte Carlo, can also be used to sample from the support more efficiently, and to obtain higher acceptance rates (see [1]).

6. References

[1] C. CHAN SHIO, "Sampling the Solutions of Differential Systems", *Ph.D. Thesis, University of Nice-Sophia Antipolis*, 2014.

[2] J.-P. COMET,, G. BERNOT, A. DAS, F. DIENER, C. MASSOT, A. CESSIEUX, "Simplified Models for the Mammalian Circadian Clock", *Procedia Computer Science*, vol. 11, p. 127-138, 2012.

[3] M. B. ELOWITZ, S. LEIBLER, "A synthetic oscillatory network of transcriptional regulators", *Nature*, vol. 403, p. 335-338, 2000.

[4] J. NOCEDAL, S. J. WRIGHT "Numerical Optimization", *Springer-Verlag, New York*, 1999.

[5] J. PRITCHARD, M. SEIELSTAD, A. PEREZ-LEZAUN, M. FELDMAN, "Population growth of human Y chromosomes: a study of Y chromosome microsatellites", *Molecular Biology and Evolution* vol. 16, p. 1791-1798, 1999.

[6] R. RAMSAY, G. HOOKER, D. CAMPBELL, J. CAO, "Parameter estimation for differential equations: a generalized smoothing approach", *Journal of the Royal Statistical Society* vol. 69, num. 5, p. 741-796, 2007.

[7] D. B. RUBIN, "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician", *The Annals of Statistics* vol. 12, p. 1151-1172, 1984.

[8] J. M. VARAH, " A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equations", *SIAM J. Sci. and Stat. Comput.* vol. 3, num. 1, p. 28-46, 1982.