Failure of maximum likelihood methods for chaotic dynamical systems

Kevin Judd
School of Mathematics and Statistics, University of Western Australia, Perth, Australia
(Received 23 June 2006; revised manuscript received 29 November 2006; published 15 March 2007)

The maximum likelihood method is a basic statistical technique for estimating parameters and variables, and is the starting point for many more sophisticated methods, like Bayesian methods. This paper shows that maximum likelihood fails to identify the true trajectory of a chaotic dynamical system, because there are trajectories that appear to be far more (infinitely more) likely than truth. This failure occurs for unbounded noise and for bounded noise when it is sufficiently large and will almost certainly have consequences for parameter estimation in such systems. The reason for the failure is rather simple; in chaotic dynamical systems there can be trajectories that are consistently closer to the observations than the true trajectory being observed, and hence their likelihood *dominates* truth. The residuals of these truth-dominating trajectories are not consistent with the noise distribution; they would typically have too small standard deviation and many outliers, and hence the situation may be remedied by using methods that examine the distribution of residuals and are not entirely maximum likelihood based.

DOI: 10.1103/PhysRevE.75.036210 PACS number(s): 05.45.−a, 02.50.Tt, 02.30.Zz

I. INTRODUCTION

A good deal of science involves solving an *inverse problem* to decide what are the most likely values of *variables* and *parameters* of a model, given observations of a system. One of the most powerful, and frequently applied, techniques for this purpose is *maximum likelihood*. If $M(\lambda)$ is a model of a system where λ is the variables and parameters to be determined, θ is the observations, and $\Pr(\theta|\lambda)$ is the probability of making the observations θ given λ, then maximum likelihood solves the inverse problem by finding the λ that maximizes $\Pr(\theta|\lambda)$. This λ is the *most likely* under the assumption that the model and distribution $\Pr(\theta|\lambda)$ are a *perfect* representation of the system and the errors of the observations.

Here we consider what happens when maximum likelihood is used to determine trajectories of chaotic dynamical systems. This is of interest in forecasting and parameter estimation of nonlinear systems [1–5], and although we will not explicitly discuss these goals, it should be obvious why our results are relevant. Put bluntly, we find that maximum likelihood cannot determine the true trajectory even when using a perfect model and unlimited data, because there exist trajectories whose likelihood dominates that of the true trajectory in the sense that these trajectories appear to be, in the limit, infinitely more likely than the true trajectory. This implies that for inverse problems involving nonlinear dynamical systems, maximum likelihood may not be reliable and may need to be supplemented or replaced by more sophisticated techniques—for example, by considering the distribution of residual errors.

The problem of determining trajectories of dynamical systems predates the maximum likelihood principle. The origin of the maximum likelihood principle is found in the work of Laplace, Legendre, Gauss, and others [6], in the guise of the least-squares method applied to the determination of astronomical orbits of two bodies.

There is a growing need to identify the trajectories of chaotic dynamical systems—for example, in weather and climate forecasting [7], in tracking tumbling ballistic missiles [8], and other applications.

Various algorithms have been proposed to identify trajectories of chaotic systems. Of course, there is the original least squares method and more recent developments in variational assimilation [9], which are only successful for short lengths of chaotic trajectories. For longer trajectories, gradient descent methods [10–18] have been successful. These do not necessarily attempt to find a maximum likelihood trajectory, only a *shadowing trajectory*, although they can be modified to do so [17]. Only recently has there been any proof that such methods actually converge to the true trajectory in the perfect model scenario, and then only for hyperbolic systems with small, bounded, additive observational noise [19]. This paper in part reveals why the small and bounded restriction is needed and what goes wrong if it is removed.

There are many filtering techniques for the estimation of the state of stochastic systems, and these have been applied to chaotic dynamical systems. These include Kalman filters, the more general Bayesian filters, various Monte Carlo simulations of Langevin equations, and so on. None of these strictly apply to deterministic chaotic systems. Furthermore, gradient descent methods appear to be more accurate than Kalman filters [20] and certainly more efficient than the Monte Carlo approaches. In any case, the problem we describe here will equally apply to these approaches if they attempt to find maximum likelihood trajectories, the problem being with the maximum likelihood principle when applied to chaotic systems, not the specific means of implementing it.

Several recent papers (cited in the second paragraph) have considered the problem of estimation of parameters of deterministic dynamical systems. This problem implicitly includes a trajectory estimation, because conceptually one can decompose the maximum likelihood principle into the likelihood of a trajectory for fixed parameter values, and maximization of this likelihood by varying the parameters. Pisarenko and Sornette [3] provide an excellent review of the maximum likelihood principle in this context, giving a clear and careful account of how various proposed methods relate to established statistical theory.

This paper has a very specific goal: to demonstrate that maximum likelihood is flawed when applied to chaotic dynamical systems under certain circumstances. In statistical terminology we show that maximum likelihood does not provide a *consistent estimator* of the true trajectory. We do not attempt to fully explore the consequences of this flaw, although there is a brief discussion at the end of the paper.

II. LIKELIHOOD

Assume we have a *perfect* model f of a discrete time deterministic dynamical system, where f is defined as a map on a bounded state space $S \subseteq \mathbb{R}^d$. [For an obscure technical reason we will assume that there exists a Borel measure μ on S with $\mu(S)=1$.] Let x_t denote the *state* of the model at time t; then, a *trajectory* of the model is a sequence of states x_t with $x_{t+1}=f(x_t)$. Let $\mathcal{T}$ denote the set of all trajectories. Assume that every observation $\theta_t \in \mathbb{R}^k$ is a random variable with a fixed known probability density $\rho(\theta_t|x_t)$ and that every observation is independent of all other observations. (We could employ weaker assumptions.) Write θ to denote the time series of observations of a *true trajectory* $x^\star \in \mathcal{T}$.

Given observations θ of $x^\star$, define

$$p(x|\theta,T) = \prod_{-T<t\leq 0} \rho(\theta_t|x_t), \tag{1}$$

which is the *likelihood* of a trajectory x given the observations θ in the time window $-T<t\leq 0$.

The question asked here is: what typically happens to $p(x|\theta,T)$ for large T? Does $p(x|\theta,T)$ concentrate on $x^\star$? To address this there is advantage in dealing with *likelihood ratios*. Suppose we know for some trajectory x that $p(x|\theta,T)>0$ for all T. Given this *reference* trajectory x, then for any other *test* trajectory y define

$$R_\theta(y,x) = \lim_{T\to\infty} \frac{p(y|\theta,T)}{p(x|\theta,T)} = \lim_{T\to\infty} \prod_{-T<t\leq 0} \frac{\rho(\theta_t|y_t)}{\rho(\theta_t|x_t)}, \tag{2}$$

assuming the limit exists, otherwise $R_\theta(y,x)$ is undefined.

III. DOMINATING TRUTH

Let $\mathcal{F}_\theta(x) \subseteq \mathcal{T}$ be the set of trajectories y for which $R_\theta(y,x)$ is defined. We say that the likelihood of x *dominates* the likelihood of y if $y \in \mathcal{F}_\theta(x)$ and $R_\theta(y,x)=0$. In other words, dominance means that as the time window of observation is extended, x appears to be far more (infinitely more) likely than y.

In this section we will show that there can exist trajectories that dominate $x^\star$—that is, appear to be far more likely than the true trajectory. We will show this using three increasingly more specific examples of one-dimensional maps, but it should be clear that the results generalize. Sections III A and III B establish the motivation and plausibility of dominance. Section III C provides a rigorous demonstration of the existence of truth-dominating trajectories in a specific class of one-dimensional maps. Sections III E and III F show numerical identifications of truth-dominating trajectories in the Henon map and their properties. Section III D is an aside on maximum likelihood trajectories and their connection with indistinguishable states [17].

It will be useful in the following to partition $\mathcal{F}_\theta(x)$ into two disjoint subsets $\mathcal{F}_\theta^0(x)$, where $R_\theta(y,x)=0$, and $\mathcal{F}_\theta^+(x)$, where $R_\theta(y,x)>0$. Note $\mathcal{F}_\theta^0(x)$ is the trajectories dominated by x.

A. Log-likelihood limits

Suppose $k=d=1$ with additive Gaussian noise with mean zero and standard deviation σ. Take $x^\star$ as the reference trajectory, let $\theta_t = x_t^\star + \delta_t$, and consider a test trajectory y and define $w_t = y_t - x_t^\star$. The δ_t represent additive observational noise and are independent of $x_t^\star$ and each other. The w_t represent the deviation of y_t from $x_t^\star$. Then

$$\frac{\rho(\theta_t|y_t)}{\rho(\theta_t|x_t^\star)} = \frac{\exp[-(\theta_t - y_t)^2/2\sigma^2]}{\exp[-(\theta_t - x_t^\star)^2/2\sigma^2]} \tag{3}$$

$$= \frac{\exp[-(\delta_t - w_t)^2/2\sigma^2]}{\exp(-\delta_t^2/2\sigma^2)} \tag{4}$$

$$= \exp[-(w_t^2 - 2\delta_t w_t)/2\sigma^2]. \tag{5}$$

Consequently, using Eq. (5), we have that

$$\log_e R_\theta(y,x^\star) = \frac{1}{2\sigma^2} \sum_{t\leq 0} (2\delta_t w_t - w_t^2), \tag{6}$$

provided $\rho(\theta|x_t^\star)>0$ for all t and the various limits and sums exist. In Eq. (6) there are four cases possible.

(i) The sum diverges to $-\infty$, implying $y \in \mathcal{F}_\theta^0(x^\star)$; that is, $x^\star$ dominates y.

(ii) The sum converges to a finite value, which implies that $y \in \mathcal{F}_\theta^+(x^\star)$. In this case neither of $x^\star$ and y are dominant. Note that if the sum is positive, then for T sufficiently large $p(y|\theta,T) > p(x^\star|\theta,T)$, and if the sum is negative, then $p(y|\theta,T) < p(x^\star|\theta,T)$.

(iii) The sum diverges to $+\infty$, which implies that $x^\star$ cannot be used as a reference trajectory. In this case $x^\star \in \mathcal{F}_\theta^0(y)$; that is, y dominates $x^\star$. [This can be seen by exchanging the roles of $x^\star$ and y in the above, which leads to case (i).]

(iv) The sum does not otherwise converge. (For example, the partial sums oscillate between two finite values.)

When and why case (iii) occurs is of most interest to us. When case (iii) holds there may be other trajectories that dominate y. Observe that the relation $2\delta_t w_t - w_t^2 > 0$ can be interpreted geometrically as saying y_t is closer to θ_t than $x_t^\star$ is. See Fig. 1, which is drawn in two dimensions, even though we only discuss one-dimensional maps. Note that y_t is inside the sphere centred on θ_t with $x_t^\star$ at its surface. The sum (6) diverges to $+\infty$ when terms corresponding to being inside these spheres outweigh the terms corresponding to being outside these spheres.

At first sight it may seem implausible that a trajectory y_t can achieve the necessary correlation with random observation variables θ_t, but we will see for chaotic dynamical systems (with generating partitions) that such correlation is easily achieved.

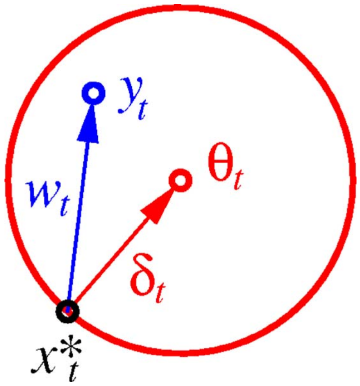

FIG. 1. (Color online) The geometrical relationship of truth $x_t^\star$, an observation θ_t, a test trajectory state y_t, and the vectors δ_t and w_t. When y_t is inside the circle, then the test trajectory state is more probable, given θ_t, than truth at this time t, or $(2\delta_t - w_t)w_t > 0$.

B. Plausible symbolic dynamics

On $[0,1)$, for $0 < \lambda < 1/2$, define a map

$$f(z) = \begin{cases} z/\lambda, & 0 \leqslant z < \lambda, \\ (z - 1 + \lambda)/\lambda, & 1 - \lambda \leqslant z < 1. \end{cases} \tag{7}$$

A middle-$(1-2\lambda)$ Cantor set is invariant under this map [21,22]. Let S be this invariant set, and consider f and S to define a dynamical system. Partition S into two components $S_0 = \{z \in S : z < 1/2\}$ and $S_1 = S \setminus S_0$. Let $s_t \in \{0,1\}$ record the sequence of visitations to S_0 and S_1 of a trajectory x. The sets S_0 and S_1 are a *generating partition* [23,24], and so the point $x_0 \in S$ is (almost surely) uniquely defined by the infinite binary sequence $(s_n)_{n=0}^{\infty}$ and it can be easily shown that

$$x_t = \sum_{n=0}^{\infty} \lambda^n (1-\lambda) s_{n+t}. \tag{8}$$

Let $s_t^\star$ be the symbol of $x_t^\star$ and define

$$s_t = \begin{cases} 1, & x_t^\star \leqslant \lambda \text{ and } \theta_t > (1+\lambda)/2, \\ 0, & x_t^\star \geqslant 1 - \lambda \text{ and } \theta_t < (1-\lambda)/2, \\ s_t^\star, & \text{otherwise.} \end{cases} \tag{9}$$

This symbol sequence defines a trajectory y_t such that if s_t and $s_t^\star$ differ, then y_t is surely closer to θ_t than $x_t^\star$. It is plausible that this y will dominate $x^\star$; however, we need to show that the likelihood gains to y by switching partitions outweighs any other variation in the likelihood it causes.

C. Likelihood growth rate with window size

To show that truth-dominating trajectories exist we demonstrate their construction for the maps (7) by making symbol changes like (9). For (7) consider the situation that leads to the key equation (6). Suppose a test trajectory y has a symbol sequence that differs from that of $x^\star$ by just one symbol at time $t=0$. It follows from Eq. (8) that

$$|w_t| = |y_t - x_t^\star| = \begin{cases} \lambda^{-t}(1-\lambda), & t \leqslant 0, \\ 0, & t > 0. \end{cases} \tag{10}$$

Using Eq. (10) in Eq. (6) gives

$$2\sigma^2 \log_e R_\theta(y, x^\star) = 2(1-\lambda)\left(\sum_{t\leqslant 0} \lambda^{-t} \delta_t \operatorname{sgn}(w_t)\right) - \lambda^2 \frac{1-\lambda}{1+\lambda}. \tag{11}$$

We want to investigate what typically happens in a situation where the symbol change puts y_0 closer to θ_0 than $x_0^\star$, but we must take into account that the symbol change may increase the distance of y_t from θ_t for $t < 0$. Under the assumption that y_0 is closer to θ_0 than $x_0^\star$, then

$$\left|\sum_{t\leqslant 0} \lambda^{-t} \delta_t \operatorname{sgn}(w_t)\right| \geqslant |\delta_0| - \sum_{t<0} \lambda^{-t} |\delta_t|. \tag{12}$$

The δ_t are random variables, so consider the expected value of this expression over noise realizations as follows. The expected value of the δ_t depends on the conditions under which we choose to change the symbol. Suppose we set a threshold $\xi > 0$, so if $|\delta_0|$ exceeds ξ with the appropriate sign, as in (9), then the symbol change is made. For a lower bound estimate (12) we can assume (without loss of generality) that $\delta_0 > \xi$, but $\delta_t \leqslant \xi$, for $t < 0$. Given the symbol change threshold ξ, we have

$$E[\delta_0 | \delta_0 > \xi] = \frac{\sigma e^{-\xi^2/2\sigma^2}}{\sqrt{2\pi}(1-p)} \tag{13}$$

and

$$E[\delta_t | \delta_t \leqslant \xi] = \frac{-\sigma e^{-\xi^2/2\sigma^2}}{\sqrt{2\pi} p}, \tag{14}$$

where $p = \Phi(\xi/\sigma)$ and Φ is the cumulative probability function of the standard normal density. Combining these two expectations (12) and (11) gives

$$E[2\sigma^2 \log_e R_\theta(y, x^\star)] \geqslant K(\xi, \sigma, \lambda) = \sqrt{\frac{2}{\pi}}\left(\frac{1-\lambda}{1-p} - \frac{\lambda}{p}\right)\sigma e^{-\xi^2/2\sigma^2} - \lambda^2 \frac{1-\lambda}{1+\lambda}. \tag{15}$$

If $K > 0$, then the likelihood of y exceeds that of $x^\star$. This can be ensured by making ξ sufficiently large, as is now shown. Since $p = \Phi(\xi/\sigma)$, then for large ξ, p tends to 1, and from asymptotic properties of the complementary error function, $1-p$ varies like $(1/\xi)e^{-\xi^2}$. Consequently, for large ξ, K varies proportional to ξ with a positive constant. Hence, for sufficiently large ξ, $K > 0$, and the likelihood of y exceeds that of $x^\star$. In fact ξ does not need to be particularly large. Using the above asymptotic approximation, for $\xi > 0$ we have $1 - p \leqslant \sqrt{2/\pi}(\sigma/\xi)e^{-\xi^2/2\sigma^2}$, also $p \geqslant 1/2$ and $e^{-\xi^2/2\sigma^2} \leqslant 1$, so $K > 0$ when

$$\xi \geqslant \frac{\lambda^2}{1+\lambda} + \frac{2\sqrt{2}\lambda\sigma}{\sqrt{\pi}(1-\lambda)}. \tag{16}$$

Since $\lambda < 1/2$, then $\xi \geqslant 2\lambda\sigma + 1/12$ is sufficient.

Now consider making multiple symbol changes by waiting at least m steps before taking advantage of another symbol change. Each additional symbol change increases the likelihood of the new trajectory y. The log-likelihood will grow linearly with window size by a rate of at least

$$K(\xi,\sigma,\lambda) \times \left(\frac{1-p}{1+m(1-p)}\right) + O(\lambda^m), \qquad (17)$$

because on average, we wait $m+1/(1-p)$ steps before making a symbol change, which when made increases the likelihood by at least the amount for the one symbol change K, defined in Eq. (15), up to terms of order λ^m. The linear growth rate implies the likelihood of y will dominate $x^\star$.

Hence, we have proven that for the shift maps with Gaussian observational noise there exist, for typical observations, trajectories that dominate truth. It should be fairly clear from following the above methodology that truth-dominating trajectories also exist for bounded noise provided the noise is sufficiently large to bridge the gap between S_0 and S_1 with sufficient frequency. In general the larger the noise, the more often one obtains advantageous symbol changes, hence the stronger the growth of the likelihood of dominating trajectories, and hence the more obvious dominating trajectories become.

The requirement that a system have a generating partition is not essential for proving the existence of truth-dominating trajectories. First note that in systems that do have generating partitions, an arbitrary partition typically provides symbol sequences that are a subshift; that is, certain symbol sequences are disallowed. Then, in the above argument, we cannot necessarily make an advantageous symbol change whenever we want. This means that we may have to search longer until we obtain an advantageous symbol change, which implies that the estimated rate of growth of the log-likelihood with window size is slower, but still linear. In practice it has been observed that many systems, with suitable partitions, have symbol sequences that are well approximated by subshifts [24–26], which suggests the above argument applies to them too. It should also be clear that the above methodology can be applied to systems with $d>1$.

D. Maximum likelihood trajectories

The previous subsection shows that truth-dominating trajectories exist. It may be asked, does there exist a *maximum likelihood trajectory* $\hat{x}$? Such a trajectory should have $\mathcal{F}_\theta(\hat{x})=S$ and $R_\theta(y,\hat{x})\leqslant 0$ for all $y\in\mathcal{F}_\theta(\hat{x})$; that is, $\hat{x}$ either dominates other trajectories or has the largest likelihood ratio when it does not dominate. It is not clear that such a $\hat{x}$ need exist, but there is reason to suspect they do for dynamical systems with generating partitions.

When there is a generating partition one might guess that the trajectory that has the same symbol sequence as θ_t is a candidate for $\hat{x}$. This is not necessarily the case. In principle, if a $\hat{x}$ exists, it could be found by a combinatorial optimization over symbol sequences, which would be a hard problem to solve. In the case of shift maps (7), however, this problem should be solvable for λ small as follows. Consider the symbol sequences satisfying

$$s_t = \begin{cases} 1, & \theta_t > (1+\lambda)/2, \\ 0, & \theta_t < (1-\lambda)/2, \\ & \text{unspecified otherwise}. \end{cases} \qquad (18)$$

A trajectory with such a symbol sequence is close to θ_t when a symbol is specified. Furthermore, by methods similar to the previous section, it can be shown that when λ is small, changing a symbol from the value specified above will decrease the likelihood. Consequently, the problem of finding $\hat{x}$ is reduced to determining the unspecified symbols for $(1+\lambda)/2\leqslant\theta_t\leqslant(1-\lambda)/2$. When λ is small there is a low probability of unspecified symbols, so they are widely separated, and specifying these symbols tends not to effect the specification of the others. Hence, the problem should be solvable sequentially. This general argument should generalize when a generating partition is available, possibly employing partition refinements and symbolic shadowing [24].

There are some properties of dominance and maximum likelihood trajectories that may be of peripheral interest. (The technical details are omitted here, but may appear somewhere more appropriate.) There is a connection between the *indistinguishable states* $H(x)$ of a trajectory x and dominance of this trajectory. (For additive Gaussian observational errors two trajectories x and y are indistinguishable if $\Sigma_{t\leqslant 0}|y_t-x_t|^2<\infty$ [17].) For mixing systems, or perhaps just systems with a positive Lyapunov exponent, and Gaussian observational errors, if $y\in\mathcal{F}_\theta^+(x)$, then θ-almost surely $H(y)\subseteq\mathcal{F}_\theta^+(x)$, and similarly for $\mathcal{F}_\theta^0(x)$. So indistinguishable states are equivalent in terms of dominance. Furthermore, if an invariant set S of a system has a generating partition and there exists a maximum likelihood trajectory with $\hat{x}_0\in S$, then for any trajectory y on S there exist trajectories $\hat{y}\in H(y)$ such that $\hat{y}_0$ is arbitrarily close to $\hat{x}_0$. In particular, this holds for the true trajectory $x^\star$. These trajectories that pass close to $\hat{x}_0$ may have small Q probability [17]; that is, they are not close to $x^\star$. The calculations of the next section illustrate this.

E. Logistic map

The construction of Sec. III C only proves the existence of truth-dominating trajectories in a class of one-dimensional shift maps. In this section we describe a suboptimal method for finding (potentially) truth-dominating trajectories by locally optimal selection of pre-images of the one-dimensional logistic map

$$f(z)=4z(1-z), \quad z\in[0,1]. \qquad (19)$$

This map has a generating partition; it is (essentially) topologically conjugate to Eq. (7) with $\lambda=1/2$.

The logistic map is two to one for every state, except $x=1/2$, so given any $y_t\neq 1/2$ there are two choices of y_{t-1}. Given observations θ of a true trajectory $x^\star$, one can work backward sequentially selecting pre-images to maximize $\rho(\theta_{t-1}|y_{t-1})$. This is suboptimal because it ignores the possibility that a pre-image selection that increases the likelihood at some t may result in later cumulative decreases that outweigh the advantage at t or preclude more advantageous changes further on. The procedure can be initialized with an

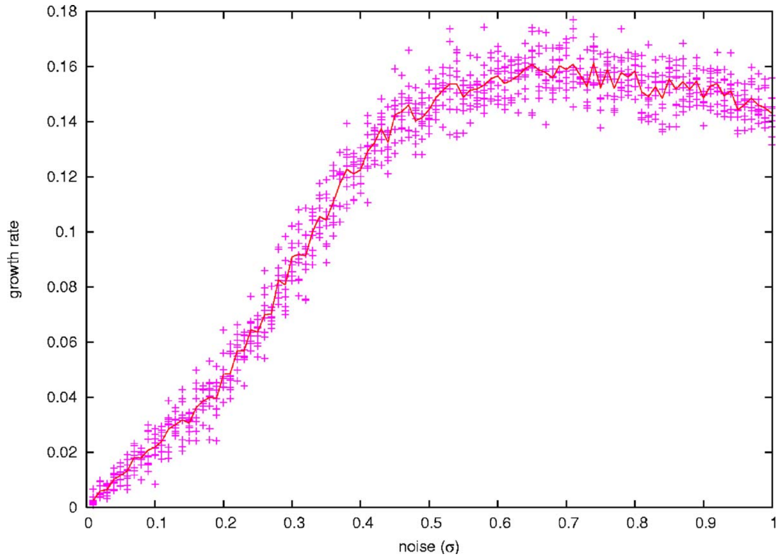

FIG. 2. (Color online) Growth rate of $\log_e R_\theta(y,x^\star)$, for different Gaussian noise $N(0,\sigma^2)$, for sequentially constructed trajectory using 3000 observations. Ten noise realizations at each level with the solid line at the mean of the realizations. True trajectory had $x_0^\star = 1/\sqrt{17}$.

arbitrary y_0; we used $y_0 = \theta_0$ in the plots, but essentially identical results were obtained for $y_0 = x_0^\star$ or arbitrary y_0.

Figure 2 shows the growth rate of $\log_e R_\theta(y,x^\star)$ computed for Gaussian noise $N(0,\sigma^2)$. From the graphs it is seen that variation of the growth rate for different observation sequences is sufficiently small given this long sample (i.e., for $\sigma > 0.1$ the mean graph is more than two standard deviations above zero) to imply that truth-dominating trajectories are definitely being identified for all but the smallest noise levels.

The construction used here could be used to find truth-dominating trajectories in other noninvertible maps. Invertible maps would require some other kind of search. For example, symbolic shadowing [24] might be adapted for this purpose. In general, in invertible maps, truth-dominating trajectories will be a species of homoclinic trajectory—that is, trajectories that repeatedly diverge from, then later return to, a neighborhood of the true trajectory's state. Consequently, one should expect to find truth-dominating trajectories near intersections of the local stable and unstable manifolds of the true trajectory [19]. Other methods of finding truth-dominating trajectories could be patching together of periodic orbits or the use of gradient descent algorithms. Whatever, the author expects that truth-dominating trajectories exist in typical chaotic systems when the observational noise is sufficiently large.

F. Residuals

Figure 3 shows the standard deviation of the residuals (difference between observations and trajectory) for both the true trajectories and the computed truth-dominating trajectories and the root-mean-square deviation of the truth-dominating trajectory from the true trajectory.

First note that the residuals of the truth-dominating are smaller than those of the true trajectory (i.e., the actual noise), and more so for larger noise levels. This is as expected because the truth-dominating trajectories will tend to be closer to the observations on average. Figure 4 shows that the residuals for the truth-dominating trajectories are not Gaussian; they are leptokurtic—that is, peaked around zero, with fatter tails than a Gaussian. This is an important observation, because it suggests that methods other than maximum likelihood that take into account the distribution of residuals may provide better methods for identifying the true trajectory.

Another interesting aspect of Fig. 3 is the root-mean-square deviation of the truth-dominating trajectory from the true trajectory. This deviation is a substantial fraction of the noise level for noise levels greater than 0.1. This implies that these truth-dominating trajectories do not look like the true trajectory a substantial fraction of the time. (They are not closely shadowing the true trajectory.) This also implies that

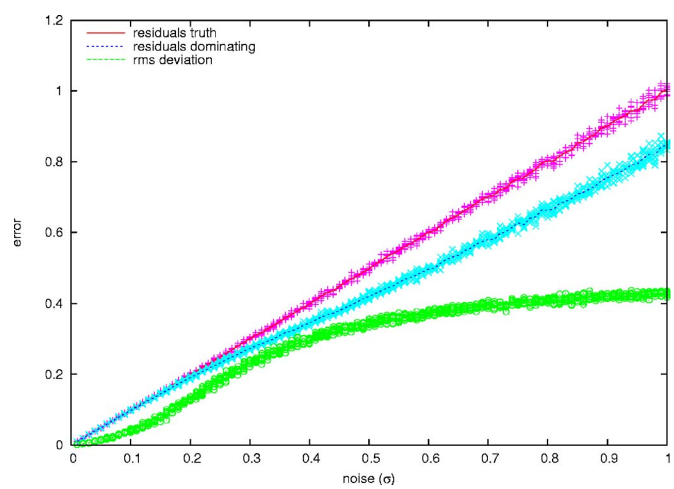

FIG. 3. (Color online) The standard deviation of the residuals of the true trajectories (+, mean as solid line, red in color, essentially the upper line of data) and the computed truth-dominating trajectories (×, mean as dashed line, blue in color, essentially the middle line of data) and the root-mean-square deviation of the truth-dominating trajectory from the true trajectory (⊙, green in color, essentially the lower line of data). Same data as used in Fig. 2. Values for ten realizations at each noise level with solid line indicating the mean of realizations.

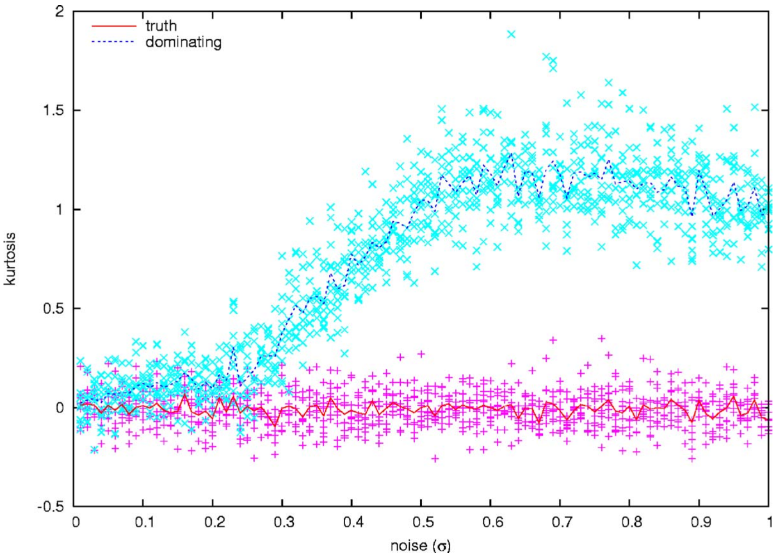

FIG. 4. (Color online) The kurtosis of the residuals of the true trajectory (+, mean as solid line, red in color, essentially lower line of data) and computed truth-dominating trajectories (×, mean as dashed line, blue in color, essentially upper line of data). Same data as used in Fig. 2. Values for ten realizations at each noise level with solid line indicating the mean of realizations.

if there is a maximum likelihood trajectory $\hat{x}$, it also differs from the true trajectory a substantial fraction of the time.

IV. CONCLUSIONS

We have defined the concept of a truth-dominating trajectory, proven that they exist in one-dimensional shift maps, given numerical evidence of them in the logistic map, and argued that they will be found in typical chaotic dynamical systems. The numerically calculated truth-dominating trajectories were significantly different from true trajectories, with deviations a substantial fraction of the noise level, and these features should be expected in general. Hence we conclude that maximum likelihood is not a consistent estimator of trajectories of chaotic dynamical systems, because it does not reliably identify the true trajectory given any amount of observations, if the noise level is sufficiently large. Dominance defeats Bayesian methods too. A sensible posterior will employ a product of $p(y|\theta,T)$ and some prior, but as T becomes large, the prior should be irrelevant.

The failure of maximum likelihood and Bayesian methods in this situation does not mean all is lost, because the fault lies with the methods and their assumptions, not statistics in general. The difficulties that arise are akin to overfitting data, where truth-dominating trajectories overfit the dynamics. In Sec. III F this fact was seen to be revealed by examining the residuals of the fitted trajectories. A true trajectory will have residuals consistent with the known observational noise distribution with probability one. Figure 4 showed that the residuals for the truth-dominating trajectories clearly are not Gaussian. This suggests a method that minimizes both the variance and kutosis of residuals might perform better at identifying the true trajectory when the observational noise is Gaussian. In general, given a model for the observational error distribution, one needs to ensure that the residuals are consistent with the noise model and, if possible, develop algorithms that use this as the criterion for finding the true trajectory, rather than simply maximizing the likelihood.

Finally, consider the question of what effect dominance might have on parameter estimation. At this stage it is an open question that needs investigation. An immediate guess might be that dominance has little effect because there are several examples in the literature of successful application of maximum likelihood techniques to parameter estimation of chaotic dynamical systems. It should be noted, however, that these consider relatively simple, lower-dimensional systems. (The author is interested in the situation for high-dimensional systems—for example, in weather and climate models.) In low-dimensional systems like the logistic map, the attractor is well sampled and has many well-defined features that are densely sampled. For example, the edges of the attractor are images of the critical point of the map. Consequently, the size of the attractor is a good indicator of the parameter value. In high-dimensional systems, the attractor is poorly sampled and not easily delineated; all points being more or less equidistant from each other. The author believes that dominance plays a more significant role in high-dimensional systems. It remains to be seen whether this is true. What effect dominance might have is currently unknown. One possible effect occurs when a parameter makes a model become more chaotic—that is, larger Lyapunov exponent or closer to being conjugate to a full shift. When estimating the parameters of this system, models with incorrect parameter values can have trajectories that match the noise better, according to maximum likelihood, than trajectories of the model with the true parameter value, which is less chaotic. Hence, maximum likelihood would give biased parameter estimates.

ACKNOWLEDGMENT

This research was supported by Australian Research Council Grant No. DP0662841.

[1] M. L. Berliner, J. Am. Stat. Assoc. **86**, 939 (1991).
[2] P. E. McSharry and L. A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).
[3] V. F. Pisarenko and D. Sornette, Phys. Rev. E **69**, 036122 (2004).
[4] D. A. Smirnov, V. S. Vlaskin, and V. I. Ponomarenko, Phys. Lett. A **336**, 448 (2005).
[5] H. U. Voss, J. Timmer, and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **14**, 1905 (2004).
[6] S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900* (Belknap Press of Harvard University Press, Cambridge, MA, 1986).
[7] T. Palmer, Rep. Prog. Phys. **63**, 71 (2000).
[8] D. C. Gompert and J. A. Isaacson (unpublished).
[9] P. Coutier, J. Meteorol. Soc. Jpn. **75**, 211 (1997).
[10] M. Davies, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 113 (1992).
[11] M. Davies, Physica D **79**, 174 (1994).
[12] C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer, Phys. Rev. Lett. **65**, 1527 (1990).
[13] S. M. Hammel, J. A. Yorke, and C. Grebogi, J. Complex. **3**, 136 (1987).
[14] C. Grebogi, S. M. Hammel, and J. A. Yorke, Bull., New Ser., Am. Math. Soc. **19**, 465 (1988).
[15] S. M. Hammel, Phys. Lett. A **148**, 421 (1990).
[16] K. Judd, Phys. Rev. E **67**, 026212 (2003).
[17] K. Judd and L. A. Smith, Physica D **151**, 125 (2001).
[18] K. Judd and L. A. Smith, Physica D **196**, 224 (2004).
[19] D. Ridout and K. Judd, Physica D **165**, 27 (2001).
[20] K. Judd, Physica D **183**, 273 (2003).
[21] R. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading, MA, 1987).
[22] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, England, 1993).
[23] M. B. Kennel and M. Buhl, Phys. Rev. Lett. **91**, 084102 (2003).
[24] Y. Hirata, K. Judd, and D. Kilminster, Phys. Rev. E **70**, 016215 (2004).
[25] G. Froyland, Nonlinearity **12**, 79 (1999).
[26] M. B. Kennel and A. I. Mees, Phys. Rev. E **66**, 056209 (2002).