2004 International Symposium on Nonlinear
Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Approximating stochastic process rates with nonlinear functions

David M. Walker† and Glenn Marion‡

†Biomathematics and Statistics Scotland, Macaulay Institute, Craigiebuckler, Aberdeen AB24 2SB
‡BioSS, University of Edinburgh, James Clerk Maxwell Building, King's Buildings, Edinburgh EH9 3JZ
Email: d.walker@bioss.ac.uk, g.marion@bioss.ac.uk

Abstract—We demonstrate how unknown process rates within a stochastic modelling framework can be approximated from time series data using polynomial functions. The problem of model selection is considered and a possible solution is suggested by adapting basis subset selection methods and description length ideas previously developed for nonlinear time series reconstruction.

1. Introduction

Stochastic modelling and simulation is an approach to simulating the behaviour of complex systems such as biological populations, disease epidemics and chemical reactions. The theory of stochastic processes is also a natural framework in which to study agent-based models – complex systems where agents interact with each other and their environment using simple local rules.

We propose to represent the rate functions in a stochastic process model by nonlinear functions. We discriminate between competing models by using a novel application of a basis selection algorithm [1] and the description length model selection criterion [2] previously developed for nonlinear time series reconstruction.

In the stochastic framework the simple rules of agent-based models are described by a sequence of events determined probabilistically. The event probabilities share the common structure

$$P\Big(\mathbf{s}(t+\delta t) = \mathbf{s}(t) + \delta\mathbf{s}\Big) = R(\mathbf{s} \to \mathbf{s} + \delta\mathbf{s}; \lambda)\delta t \qquad (1)$$

where $R(\mathbf{s} \to \mathbf{s} + \delta\mathbf{s}; \lambda)$ represents the rate of an event causing the change $\delta\mathbf{s}$ in state space and λ represents parameters of the stochastic process. More generally, the parameters can depend on the system state or the time. We shall restrict our discussion and studies to constant parameterizations.

A given realization of a Markov process is

$$z = \{\mathbf{s}(t_1), \ldots, \mathbf{s}(t_N)\} \qquad (2)$$

where $t_1, \ldots, t_n$ denote the event times and $\mathbf{s}(t)$ denotes the state of the system immediately prior to the event at time t.

The model is event driven so let n index these events and denote the event type occurring at time t_n by $E(n)$. If all event types are visible and if the process is monitored continuously we observe all N events of the realization. If we denote the parameters of the process by λ then for the system in state $\mathbf{s}(t)$ for $t \in (t_0, t_1)$ the likelihood $L(\lambda, D)$ may be written

$$L(\lambda, D) = \prod_{n=1}^{N} r_{E(n)} e^{-(t_n - t_{n-1}) \sum_{j=1}^{m} r_j(\mathbf{s}(n))} \qquad (3)$$

where, in this case, the observations D are equivalent to the complete realization S, the state of the system for all $t \in (t_0, t_N)$. The term $r_{E(n)}$ corresponds to the probability of event n occurring at time n whilst the exponential term is the probability that nothing happens between event $n-1$ and event n. The total event rate is $R(\mathbf{s}(t)) = \sum_{j=1}^{m} r_j(\mathbf{s}(t))$ for m possible transition rates.

When the parameter space is low dimensional, and the data is a complete realization, it is straightforward to obtain maximum likelihood estimates of the parameters by applying standard optimization routines. In this paper we do not consider incomplete realization data concentrating only on developing methods of model selection for approximating the process rates. The calculation of the likelihood is much harder in the case of missing data and new methods must be developed. Nonetheless it is anticipated that much of the framework for model selection presented here will be applicable when only incomplete observations are available. An alternative paradigm such as Bayesian statistics and Markov chain Monte Carlo (MCMC) methods may also provide a solution. Indeed we have applied and developed algorithms for reconstructing consistent realizations of processes to be used in MCMC estimation of stochastic spatial agricultural system models [3, 4]. The Bayesian approach to model selection in classification and regression problems is discussed in the monograph [5].

2. Nonlinear function approximations

We propose to represent the (unknown) rate functions in a stochastic process model by nonlinear functions. In this paper we will consider polynomial basis functions but other representations are possible.

If the state of the system at the time of event n is denoted by $\mathbf{s}(n)$ then a polynomial approximation to a process rate is

$$R(\mathbf{s} \to \mathbf{s} + \delta\mathbf{s}) = \mathbf{A} + \mathbf{B}\mathbf{s}(n) + \mathbf{C}\mathbf{s}(n)\mathbf{s}^T(n) + \cdots \qquad (4)$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$,... are constant tensors of appropriate rank.

The values of the parameters can be estimated by minimizing the negative log likelihood where the rate functions and the event probabilities in (3) are replaced by the appropriate polynomial representations (4).

Crucial problems of using nonlinear basis function and polynomial approximations in particular, is the bias-variance problem and the curse of dimensionality; models with a large number of parameters are difficult to estimate and don't generalize well. We are therefore faced with a model selection problem, namely, given a chosen model class how do we choose an appropriate selection of basis functions?

3. Selection Algorithm

In this paper we adapt the selection algorithm developed by Judd and Mees [1] for selecting subsets of basis functions of nonlinear basis function models under Gaussian noise assumptions for time series model reconstruction. The sensitivity analysis results of optimization theory are applied to the likelihood in order to determine which model basis function should be selected for removal or inclusion. This step is computationally more complicated in our present situation due to the non-Gaussian likelihood but the arguments carry through in a similar manner.

Following Judd and Mees, the problem of selecting which basis function to add to the current model of the stochastic rate approximation can be written as

$$\text{minimize } -\log L(\lambda) \text{ subject to } N(\lambda) = k, \qquad (5)$$

where we have dropped the dependence on data in (3) for clarity in exposition. The constraint $N(\lambda) = k$ corresponds to the number of terms in the model. Setting $B = \{j : \lambda_j \neq 0\}$, so that $N(\lambda) = |B|$, we can use sensitivity analysis to see the effect of changing the size of B. We write the constraint as

$$\lambda_j = u_j, \quad j \notin B \qquad (6)$$

where $\mathbf{u} = \mathbf{0}$ but are kept as parameters. The Lagrangian [6] for (5) with (6) is

$$L(\lambda, \mu) = -\log L(\lambda) + \mu^T(\mathbf{u} - \lambda) \qquad (7)$$

where μ are dual variables. The Kuhn-Tucker conditions give rise to

$$\begin{aligned} \mu &= \nabla_\lambda(-\log L(\lambda)) \\ \mathbf{u} - \lambda &= 0 \end{aligned}$$

Since μ is the dual variable corresponding to constraint (6) it is the sensitivity to changes in $\mathbf{u}$ at optimality, and therefore the largest element of μ in absolute values should be added to the basis to give the greatest marginal improvement in cost. This gives a prescription for adding basis function to extend the model.

We can find a prescription for removing an existing basis function from the model by considering the (Lagrangian) dual problem. That is,

$$\begin{aligned} \text{maximize} \quad & -\log L(\lambda) + \mu^T(\mathbf{u} - \lambda) \\ \text{subject to} \quad & \mu_j = w_j, \quad j \in B \end{aligned}$$

where $\mathbf{w} = \mathbf{0}$ but are kept as parameters and $\lambda = \lambda(\mu)$. If we set $\mathbf{u} = \mathbf{0}$ immediately then the Kuhn-Tucker conditions show that ν – the Lagrange multiplier for this problem – is a dual variable to λ and so selecting the smallest existing λ_j in absolute value as the variable to remove from the basis will do the least damage to the cost.

The above information gives a means by which an iterative scheme for expanding and shrinking model size can operate. Following [1] the 'best' model over all model sizes can be chosen as the one which minimizes a description length criterion to be described in the next section.

4. Description Length

The criterion we use is again motivated by the methods and results of Judd and Mees [1] who applied the minimum description length (MDL) results of Rissanen [2] to discriminate between nonlinear basis function models of different size.

Description length can be used to discriminate between models of different size for the same data set by comparing the cost of describing the data in terms of code length. The data itself has a certain code length necessary for its description, i.e., the cost of representing the data using floating point representations. Alternatively, one can consider a model describing the data and calculate the cost of representing the prediction errors of the model plus the code length necessary to represent the model, or its parameters, at a certain precision. We call the code length of the model plus data (errors) the description length. The model with the minimum description length is chosen as the 'best' model.

The total description length for a realization z is

$$C(z, \bar{\lambda}) = C(z|\bar{\lambda}) + C(\bar{\lambda}) \qquad (8)$$

where the data code length can be approximated by [2]

$$C(z|\bar{\lambda}) = -\log P(z|\bar{\lambda})) \approx -\log L(\bar{\lambda}, z) \qquad (9)$$

the negative log likelihood of the data. The code length needed to specify the parameters $\bar{\lambda}$ is [1]

$$C(\bar{\lambda}) \approx \sum_{j=1}^{k} \log \frac{\gamma}{\delta_j} \qquad (10)$$

γ is the number of factors of 2 required in the exponent of a floating-point representation of a parameter. Following [1] we set $\gamma = 32$ in our experiments.

We can bound the description length by considering the maximum likelihood parameter values $\hat{\lambda}$ so that

$$C(z|\bar{\lambda}) \leq C(z|\hat{\lambda}) + \frac{1}{2}\delta^T Q \delta \tag{11}$$

where $Q = D_{\lambda\lambda} C(z|\hat{\lambda})$. Therefore,

$$C(z, \bar{\lambda}) \leq C(z|\hat{\lambda}) + \frac{1}{2}\delta^T Q \delta + k \log \gamma - \sum_{j=1}^{k} \log \delta_j \tag{12}$$

The right hand side can be minimized with respect to δ and the optimal precisions $\hat{\delta}$ are the solution to

$$(Q\delta)_j = 1/\delta_j \tag{13}$$

The approximate description length of a model with k parameters can then be written as

$$S_k(z) = -\log L(\hat{\lambda}, z) + (\frac{1}{2} + \log \gamma)k - \sum_{j=1}^{k} \log \hat{\delta}_j \tag{14}$$

For each model obtained using the selection methods of the previous section we calculate $S_k(z)$ and rank the various models. The model with the minimum $S_k(z)$ is chosen as the optimal model.

Asymptotically, the description length criterion can be simplified to the form [2]

$$M_k(z) = -\log L(\hat{\lambda}, z) + \frac{k}{2} \log N + O(k) \tag{15}$$

where N is the number of data. A related criterion for model selection is given by Akaike's AIC which can be written as [2]

$$A_k(z) = -2 \log L(\hat{\lambda}, z) + 2k \tag{16}$$

5. Examples

Logistic birth-death population model The logistic model describes general population growth in the absence of immigration and emigration. That is, only birth and death processes are considered. The birth and death rates are given by polynomial functions of the population, namely

$$\begin{aligned} B[N(t)] &= N(t)(a_1 - b_1 N(t)) \\ D[N(t)] &= N(t)(a_2 + b_2 N(t)) \end{aligned}$$

The state space is one dimensional with the state at time t given by $N(t)$. A typical simulation [7] with $N(0) = 1$, $a_1 = 2.2$, $a_2 = 0.2$, $b_1 = b_2 = 0.1$ and $t_{max} = 15$ is shown in Figure 1.

We assume the process rates can be represented by polynomial functions of maximum order 5 (c.f. order 2 in model used to generate data) and apply the methods outlined above. In Figure 2 we show the values of our model

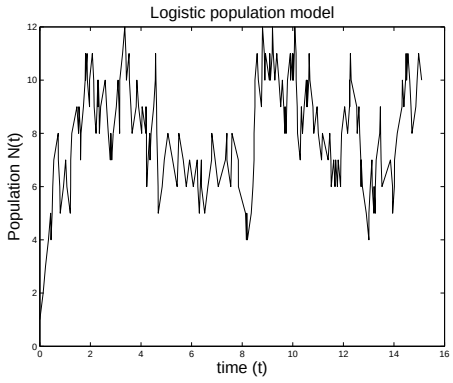

Figure 1: Typical simulation of the logistic birth-death population model.

selection criteria in this case. We see that the best model suggested by MDL is a two parameter model which corresponds to birth and death process rates linear in the population size. This is different from the known representations of the process rates but by examining Figure 1 we can see that the realization reaches a "stochastic equilibrium" quite quickly and so for a finite observation sequence it is not too dispiriting that the model selection methods favour simple model representations. For the data shown in Figure 1 several runs of the selection algorithm with different two parameter initial models were carried out which indicates the variability in the selection criteria as can be seen in Figure 2.

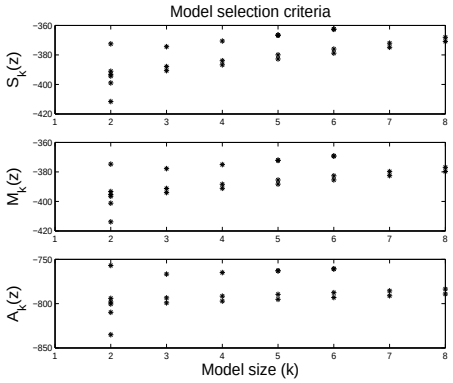

Figure 2: The top figure shows the results of using (14), the middle figure shows the results of using the asymptotic relation and the lower figure is a plot of Akaike's relationship.

Reaction time-lag birth-death population models There is often a time-lag between cause and effect; plants take time to mature before producing seeds for example. Population models which depend on the population size at the current time may not capture such behaviour, and so it is necessary to build in a natural time-lag into the mod-

els. The introduction of a time-delay vastly increases the dimension of the process and can produce quite complex dynamics as a result.

An example of a birth-death population model with delayed process rates is

$$\begin{aligned} B[N(t)] &= N(t) \\ D[N(t)] &= N(t)N(t-t_D)b_2 \end{aligned}$$

A lengthy simulation [7] with $N(0) = 50$, $t_D = 1.8$, $b_2 = 1/50$ and time step $h = 0.001$ is shown in Figure 3. The population is output every $t = 0.001s$ and recorded when the population changes by ± 1.

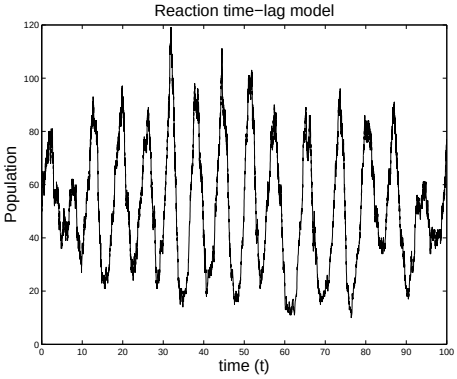

Figure 3: A typical simulation of the time-lag birth-death population model.

We attempt to reconstruct the behaviour of the population from an observation of a realization of the above stochastic time-delay model. We will consider polynomial representations of maximum order 5 of the birth and death process rates and use $\mathbf{s}(n) = (N(n), N(n-150))$ as the system state. We note that as yet we have no diagnostic, or even a prescription, for determining appropriate equivalents to the nonlinear time series reconstruction concepts of embedding dimension and time-delay lag for this application. The lag of 150 was chosen as $1/4$ of the approximate dominant period in the stochastic time series shown in Figure 3.

In Figure 4 we show the results of a run of the estimation and selection algorithm. The optimal model suggested is a two parameter model where the birth and death process rates are linear in the population. There is, however, variability in the criteria calculations over many runs with different initial model representations. For example, starting with linear models with basis term $N(n-150)$ an 11 parameter model is suggested with the birth rate a 5^{th} order polynomial.

6. Conclusion

We have introduced and demonstrated how process rates within a stochastic framework can be approximated using

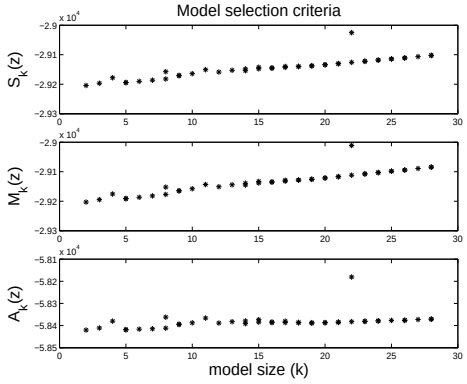

Figure 4: The top figure shows the results of using (14), the middle figure shows the results of using the asymptotic relation and the lower figure is a plot of Akaike's relationship.

polynomial functions. The problem of model selection was addressed by adapting methods used in nonlinear time series reconstruction. The scope for modelling systems using stochastic frameworks is wide and we believe the methods introduced here can aid their diverse application.

Acknowledgments

This work was funded by SEERAD.

References

[1] K. Judd and A. Mees, "On selecting models for nonlinear time series," *PHYSICA D*, Vol.82, pp.426–444, 1995.

[2] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, 1989.

[3] D. M. Walker, G. Marion, D. L. Swain, M. R. Hutchings, M. Friend and D. A. Elston, "Estimating the parameters of a stochastic grazing model," *Proceedings of the Thirty-fifth Meeting of the Agricultural Research Modellers' Group*, 2003.

[4] D. M. Walker, G. Marion, D. L. Swain and M. R. Hutchings, "Agent-based modelling with applications," in preparation, 2004.

[5] D. G. T. Denison, C. C. Holmes, B. K. Mallick and A. F. M. Smith, *Bayesian methods for nonlinear classification and regression*, Wiley, 2002.

[6] M. Parlar, *Interactive Operations Research with Maple*, Birkhauser, 2000.

[7] E. Renshaw, *Modelling Biological Populations in Space and Time*, Cambridge, 1991.