Mathematical Biosciences 243 (2013) 28–45

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

When can a deterministic model of a population system reveal what will happen on average?

Leif Gustafsson *, Mikael Sternad

Signals and Systems, Department of Engineering and Sciences, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form 4 January 2013
Accepted 11 January 2013
Available online 31 January 2013

Keywords:
Comparison of models
Consistent results
Poisson Simulation
Population model
Simulation methodology
Unbiased results

ABSTRACT

A dynamic population system is often modelled by a deterministic difference equation model to obtain average estimates. However, there is a risk of the results being distorted because unexplained (random) variations are left out and because entities in the population are described by continuous quantities of an infinitely divisible population so that irregularly occurring events are described by smooth flows.

These distortions have many aspects that cannot be understood by only regarding a deterministic approach. However, the reasons why a deterministic model may behave differently and produce biased results become visible when the deterministic model is compared with a stochastic model of the same structure.

This paper focuses first on *demographic stochasticity*, i.e. stochasticity that refers to random variations in the occurrence of events affecting the state of an individual, and investigates the consequences of omitting this by deterministic modelling. These investigations reveal that bias may be strongly influenced by the type of *question to be answered* and by the *stopping criterion* ending the analysis or simulation run. Two cases are identified where deterministic models produce unbiased state variables: (1) Dynamic systems with stable local linear dynamics produce unbiased state variables asymptotically, in the limit of large flows; and (2) linear dynamic systems produce unbiased state variables as long as all state variables remain non-negative in both the deterministic and the stochastic models. Both cases also require the question under study to be compatible with a solution over a fixed time interval.

Stochastic variability of initial values *between* simulation runs because of uncertainty or lack of information about the initial situation is denoted *initial value stochasticity*. Elimination of initial value stochasticity causes bias unless the model is linear. It may also considerably enlarge bias from other sources.

Unknown or unexplained variations from the environment (i.e. from outside the borders of the studied system) enter the model in the form of stochastic parameters. The omission of this *environmental stochasticity* almost always creates biased state variables.

Finally, even when a deterministic model produces unbiased state variables, the results will be biased if the output functions are not linear functions of the state variables.

1. Introduction

This paper examines the mathematical conditions for a deterministic difference equation model to produce unbiased results when used for population studies.

A *population* consists of entities that may have attributes of different kinds. These entities may perform different actions and they may interact with other entities and with the environment. A *population model* should therefore capture the nature of the population system under study as it evolves *dynamically* over *continuous time*, focusing on the aspects specified by the purpose of the study. Population models constitute a class of models with a *non-negative integer number of discrete entities* such as plants, animals, patients, vehicles, molecules, atoms, data packets, etc. Such models are fundamental in ecology, epidemiology, demography and queuing systems. They are also used in physics, chemistry, molecular biology, traffic planning, production planning and many other fields.

1.1. Stochastic population models – representation and time handling

In this paper the behaviour of stochastic population models is used as a reference in comparisons with that of deterministic models.

All irregularities from the behaviour of a real population, its environment and the system conditions at a given (initial) point in time are rarely known and cannot be modelled in detail. Instead, irregularities are often specified in probabilistic terms, leading to

* Corresponding author. Tel.: +46 18 500061.
E-mail address: Leif.Gunnar.Gustafsson@gmail.com (L. Gustafsson).

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.01.006

stochastic population models. A non-trivial question is then how such a probabilistic description of a dynamic population can, and should, be converted into an executable simulation model.

Three types of stochasticity may be implemented in a population model, namely demographic stochasticity, environmental stochasticity and initial value stochasticity. *Demographic stochasticity* [1,2] is the randomness about 'when the next event such as birth, death, migration or other change of state (or attribute) of an entity will occur'. Stochastic variability of initial values *between* simulation runs because of uncertainty or lack of information about the initial situation is denoted *initial value stochasticity*. This is conceptually related to demographic stocasticity, but will only intervene at the initial point of the study, while *environmental stochasticity* [1,2] comes from external, partly unknown factors such as temperature, precipitation, supply of food, etc. that can vary *within* the simulation run.

In a recent investigation of the foundations of stochastic population modelling and simulation [3], it was found that demographic stochasticity can be correctly handled by either of three different types of *representation*:

- First, the population model can use an *entity-based representation* where each entity is individually described with its attributes and conditional behaviour and can be individually monitored. This micro representation is used in e.g. Discrete Event Simulation and Individually Based Simulation.
- The second option is to use a *compartment-based representation* where all entities in the same situation (i.e. having the same set of attribute values) are located in the same compartment. Only the number of entities in each compartment is then recorded. This macro representation is used in e.g. compartment models, which can be numerically handled using tools from Continuous System Simulation.
- Third, a *state-based representation* exists, where every situation that the entire model can be in has to be represented as a unique separate state. This representation is used for e.g. Markov models and often leads to a model of gigantic size.

The *time handling*, i.e. the updating of the (demographic) stochastic model over time, is based on irregularly occurring *events* that will change the attribute values. The time handling can be performed in three different ways, irrespective of the representation used. The first way is to increment time in such small time-steps that zero or at most one event may occur during a time-step (using Bernoulli statistics). The second way is to use larger time-steps allowing several events per time-step (using Poisson statistics), and the third is to 'jump' to the point in time for the next event (using exponential statistics).

It has been shown [3] that irrespective of the representation (entity, compartment or state) and the time handling method selected, stochastic population models will produce mutually *consistent* (contradiction-free) results, provided that some simple rules are observed during the model building process.

1.2. Deterministic models of a population

Historically, there is a long tradition of using *deterministic* differential or difference equation models for analysis or simulation of populations.

Using a deterministic model simplifies model building. There is an overwhelming number of studies where populations are modelled in this way, using continuous state variables and ignoring the effects of stochasticity. Even in textbooks on e.g. epidemiology, biology and ecology [4–6] and in more general textbooks about dynamic systems [7,8], the presentations are often entirely based on deterministic models, or at least lack demographic stocasticity. It is then taken for granted that this is adequate and no concern is expressed about the validity.

On the other hand, there are early examples of understanding that stochastic and deterministic models of the same population system may produce profoundly different results. For example, in 1926 A.G. McKendrick [9] published a stochastic treatment of an epidemic process and in 1960 M.S. Bartlett [10] published his book '*Stochastic Population Models in Ecology and Epidemiology*' in which the necessity of stochastic population models was emphasised and different ways to construct them were discussed. Many, more recent, textbooks present both deterministic and stochastic modelling and stress the relative advantages of both types [11–15]. However, a general discussion of the circumstances where deterministic models are appropriate is lacking: "Full conditions for agreements between deterministic and expected stochastic solutions are at present unknown." [13]. Some studies have compared the results of different approaches to specific problems [16–22], but in general this is not done.

1.3. Objective and overview of the paper

"Understanding the dynamics of stochastic populations, and how they deviate from the deterministic ideal, is being viewed with increasing importance by ecologists and epidemiologists" [22].

It is not unusual that modellers assume that models with large numbers (or large flow rates) or that linear models will be sufficient to secure unbiased average results when stochastics are not included. Still worse, working within the deterministic tradition, this issue may not be raised at all.

Since entity-based, compartment-based and state-based models are three ways to produce a consistent realisation of a conceptual population model with demographic stochasticity [3,23], they provide a gold standard when compared with a simplified deterministic difference equation model.

The objective of this paper is to investigate when a deterministic model can reproduce the average quantities[1] of a corresponding stochastic population model and to demonstrate and explain what might otherwise happen. For this purpose, we require the quantities to be *unbiased*, in the sense that:

$$\textit{deterministic quantity} = E[\textit{stochastic quantity}],$$

where $E[\cdot]$ denotes an ensemble average for a corresponding stochastic model. In the following, this property is evaluated analytically when possible. Otherwise, a check is made of whether a quantity from the deterministic model falls within the 95% confidence interval of that from the stochastic model, based on 10,000 simulation runs.

A fundamental restriction implied by the use of a deterministic model is that all information about variations, confidence intervals, risks, extremes, correlations, etc. is lost and only deterministic results such as point estimates can be obtained instead of probability distribution functions.

Nevertheless, a deterministic model of a population may be useful in several circumstances if it can produce outcomes that are close to the average of the outcomes that would be produced by a corresponding stochastic model:

- One case is when average results are the main outcome of interest in a study. The deterministic model could then produce these average results in just one simulation run, instead of averaging over many simulation runs of a stochastic model.

[1] We need to distinguish between bias in the state variables and in the results (outcomes of the study). 'Quantity' is used to include both 'state variable' and 'result'.