Mathematical Biosciences 267 (2015) 97–108

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Probability bounds analysis for nonlinear population ecology models

Joshua A. Enszer[1], D. Andrei Măceș, Mark A. Stadtherr*

Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Received 31 January 2015
Revised 15 June 2015
Accepted 17 June 2015
Available online 4 July 2015

Keywords:
Parameter uncertainty
Probability distributions
Population ecology
Food webs
Nonlinear dynamics
Ionic liquids

ABSTRACT

Mathematical models in population ecology often involve parameters that are empirically determined and inherently uncertain, with probability distributions for the uncertainties not known precisely. Propagating such imprecise uncertainties rigorously through a model to determine their effect on model outputs can be a challenging problem. We illustrate here a method for the direct propagation of uncertainties represented by probability bounds though nonlinear, continuous-time, dynamic models in population ecology. This makes it possible to determine rigorous bounds on the probability that some specified outcome for a population is achieved, which can be a core problem in ecosystem modeling for risk assessment and management. Results can be obtained at a computational cost that is considerably less than that required by statistical sampling methods such as Monte Carlo analysis. The method is demonstrated using three example systems, with focus on a model of an experimental aquatic food web subject to the effects of contamination by ionic liquids, a new class of potentially important industrial chemicals.

1. Introduction

Mathematical models are often the only resource available to predict the effects of anthropogenic influence on ecological systems. Limited physical experiments can possibly isolate and estimate the interactions between a subset of species in an ecosystem, or determine the effects of a change to the environment (e.g., a change in some resource or the introduction of a new resource, predator, or contaminant). However, it is difficult to replicate many such interactions or changes with physical experiments.

It can also be challenging to develop and effectively use mathematical models of ecosystems, particularly in the presence of uncertainty. The importance of dealing with the many potential sources of uncertainty in developing and using population ecology models is well known [e.g., 1–4]. Our focus here is on those types of uncertainty (e.g., measurement error, natural variation) that may manifest themselves as uncertainties in model parameters. Given some quantitative description of the parameter uncertainty, such as an interval or a probability distribution, the general goal of uncertainty analysis is to quantify the effect of such uncertainty on the model outputs, or, in other words, to "propagate" the parameter uncertainty to the outputs. For relatively simple static or algebraic models, this might be done directly, perhaps using interval arithmetic or appropriate convolutions of probability distributions. For more complex or dynamic models, this is widely done using various sampling methods (e.g., Monte Carlo) in which model outputs are computed repeatedly at many different samples of the parameter values, with samples taken based on a specified probability distribution, if available.

In the presence of multiple types of uncertainty, it may be appropriate to describe the parameter uncertainty using probability bounds [3,5,6]. In this case, probability distributions are not known precisely but instead bounds on the cumulative probability distributions are given, thus effectively combining the ideas of intervals and probability distributions. For example, probability bounds may be a useful treatment of uncertainty when both measurement error (often represented by "error bars", i.e., intervals) and natural variability (often represented by probability distributions) are present. When probability bounds are used, direct propagation of the uncertainty is again possible for reasonably simple static or algebraic models, and there is software available for this purpose [6]. For more complex or dynamic models, sampling methods can again be used; however, this is now a second-order (or two-dimensional) process [7], in which first a sample of the probability distribution for the parameters is taken from within their given probability bounds, and then this probability distribution is used to sample the parameter values. Such a nested sampling procedure can become quite expensive computationally.

We illustrate here a method for the direct propagation of uncertainties represented by probability bounds through nonlinear, continuous-time, dynamic models in population ecology. Uncertainties represented by simple intervals or probability distributions can also handled, as special cases, using this approach. No sampling is

* Corresponding author. Tel.: +1 574 631 9318; fax: +1 574 631 8366.
E-mail address: markst@nd.edu (M.A. Stadtherr).
[1] Current address: Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716, USA.

http://dx.doi.org/10.1016/j.mbs.2015.06.012
0025-5564/

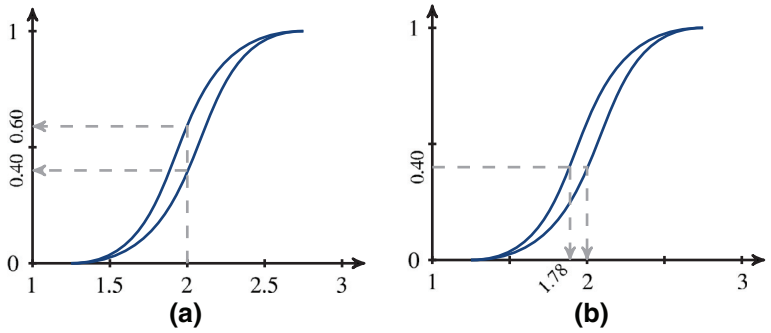

Fig. 1. Interpretation of a p-box PB(x). (a) The probability that $x \leq 2$ is bounded by the interval [0.40, 0.60]. (b) The 40th percentile value of x is bounded by the interval [1.78, 2.00].

required, and computed bounds on outputs are mathematically and computationally rigorous. This approach was originally developed [8] for applications in chemical process reaction engineering.

This paper is organized as follows. In the next section, we will provide some brief background on the key mathematical tools used, in particular intervals, and their extension to probability boxes, and Taylor models. Then we will provide a concise mathematical statement of the general problem to be solved, followed by a summary of the solution methods used. We will then demonstrate these methods using three example systems, with focus on a model of an experimental aquatic food web subject to the effects of chemical contamination.

2. Background

2.1. Interval analysis

One simple way of representing uncertainty in a model parameter is to treat it as an interval. This is appropriate if upper and lower bounds are known, but there is no information about a probability distribution. Formally, we define a real interval X as the set of real numbers between (and including) a specified lower bound (denoted by $\underline{X}$) and upper bound (denoted by $\overline{X}$). That is, $X = [\underline{X}, \overline{X}] = \{x \in \Re \mid \underline{X} \leq x \leq \overline{X}\}$. A real interval vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ has n real intervals as components and can be regarded as an n-dimensional rectangle or box. Interval matrices are similarly defined. Arithmetic on intervals is defined according to X op $Y = \{x$ op $y \mid x \in X, y \in Y\}$, op $\in \{+, -, \times, \div\}$. Division in the case of $0 \in Y$ is allowed only in extensions of interval arithmetic [9]. Interval versions of the elementary functions can be similarly defined. Interval computations are implemented with outward rounding (lower bound rounded down, upper bound rounded up). Thus, interval computations can be used to obtain rigorously guaranteed bounds on function ranges, and play a key role in the verified (or validated) numerical solution of a variety of problems in science and engineering [10].

For a real function $f(\boldsymbol{x})$ of n variables, the interval extension $F(\boldsymbol{X})$ provides bounds on the range of $f(\boldsymbol{x})$ for $\boldsymbol{x} \in \boldsymbol{X}$. That is, $\{f(\boldsymbol{x}) | \boldsymbol{x} \in \boldsymbol{X}\} \subseteq F(\boldsymbol{X})$. However, while these bounds are guaranteed, they are not necessarily tight. If $F(\boldsymbol{X})$ is computed using interval arithmetic (by replacing $\boldsymbol{x}$ with $\boldsymbol{X}$ in the expression for $f(\boldsymbol{x})$), and if any variable occurs more than once in this expression, then the function range may be overestimated due to the "dependency" problem. This occurs because, in interval arithmetic, separate occurrences of the same variable are not recognized as dependent. Another potential source of overestimation (lower bounds too low, upper bounds too high) in the use of interval methods is the "wrapping" effect [11]. This occurs when a multidimensional interval is used to enclose (wrap) a set of results that is not an interval. If this type of overestimation is propagated, say from step to step in an integration method for ordinary differential equations (ODEs), it can lead quickly to the loss of a meaningful enclosure. Historically, the issues of wrapping and dependency have resulted in interval methods acquiring a reputation for producing overly loose and conservative bounds. However, current interval methods, including the use of techniques such as Taylor models, as discussed below, can often yield rigorous bounds with very little overestimation. Several good introductions to interval analysis, as well as interval arithmetic and other aspects of computing with intervals, are available [9,10,12–15].

2.2. Probability boxes (P-boxes)

An interval gives an upper and lower bound only, and provides no knowledge about the distribution of uncertainties. If some (but not exact) knowledge about the distribution is available, then this can be is captured by using "probability boxes" (p-boxes), which provide interval-like bounds on the cumulative distribution function (CDF) [6,8,16,17]. Intervals and exact CDFs represent special cases of the more general concept of the p-box.

For some quantity (variable or parameter) x, we define the CDF $F_x(z)$ as giving the probability that $x \leq z$. A p-box for x, denoted $\mathrm{PB}(x) = (L_x, R_x)$, is the set of all such CDFs enclosed by two bounding CDFs $L_x(z)$ and $R_x(z)$ with finite support. That is, $\mathrm{PB}(x) = (L_x, R_x) = \{F_x(z) \mid L_x(z) \geq F_x(z) \geq R_x(z)\}$. For a given value of z, the left bounding function $L_x(z)$ of the p-box gives the upper bound on the probability that $x \leq z$ and the right bounding function $R_x(z)$ gives the lower bound on this probability. This is shown for an example p-box in Fig. 1(a), which is marked to indicate that, for this p-box, the probability that $x \leq 2$ is bounded by the interval [0.4, 0.6]. Conversely, for a given value of the cumulative probability, $L_x(z)$ and $R_x(z)$ provide lower and upper bounds on the values of x for which this probability is possible. For the case of the p-box in Fig. 1(b), this shows that the 40th percentile value of x is bounded by the interval [1.78, 2]. The bounding functions in Fig. 1 are (truncated) Gaussian CDFs. However, the p-box encloses both Gaussian and non-Gaussian CDFs.

Williamson and Downs [5] have presented methods for rigorously bounding the results of arithmetic (or other) operations on random variables when only their bounding distributions are known. This can be done without assuming any information about possible correlation between the operands. It can also be done for the cases that the operands are independent, or that they are the same, such as in a polynomial or other expression with a repeated operand. In general, these methods are implemented numerically, and use piecewise-constant discretizations of the bounding distributions. For the discretization, the p-box bounds are enclosed using an ordered set of d intervals, each representing a probability range of equal weight $1/d$. Subsequent operations are then done on these intervals using interval arithmetic. A detailed example of an arithmetic operation on two p-boxes is given by Enszer et al. [8], who also demonstrate how the dependency and wrapping issues extend from interval operations to p-box operations. Analogous procedures can be used to determine probability bounds on the results of other functions (e.g., logarithm, integral powers, polynomial, etc.) Obviously, a tighter enclosure of a p-box can be obtained using a finer discretization. Unless noted