Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 5090 – 5102
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Robustness of the estimators of transition rates for size-classified matrix models

M. Zetlaoui[a], N. Picard[b], A. Bar-Hen[c,*]

[a]Institut National Agronomique Paris-Grignon, 16 rue Claude Bernard, 75231 Paris Cedex 5, France
[b]Cirad, Forest Department, Campus de Baillarguet, TA 10/D, 34398 Montpellier Cedex 5, France
[c]INA-PG OMIP, 16, rue Claude Bernard, F-75231 Paris cedex 5 & Université Paris 13, SMBH-Lim & Bio, 74, rue Marcel Cachin, 93017 Bobigny Cedex, France

Available online 15 August 2006

Abstract

Matrix models are often used to model the dynamics of age-structured or size-structured populations. The Usher model is an important particular case that relies on the following hypothesis: between time steps t and $t+1$, individuals either remain in the same class, move up to the following class, or die. There are then two ways of handling data that do not meet this condition: either remove them prior to data analysis or rectify them. These two ways correspond to two estimators of transition parameters. The former, which corresponds to the classical estimator, is obtained from the latter by a data trimming. The two estimators of transition parameters are compared on the basis of their robustness in order to obtain a criterion of choice between the two estimators. The influence curve of both estimators is first computed, then their gross sensitivity and their asymptotic variance. The untrimmed estimator is more robust than the classical one. Its asymptotic variance can be lower or greater than that of the classical estimator depending on the boundaries used for data trimming. The results are applied to a tropical rain forest in French Guiana, with a discussion on the role of the class width.

Keywords: Influence curve; Matrix model; Parameter estimation; Robustness; Usher model

1. Introduction

Models of population dynamics are an important tool in many ecological studies. They are used to mimic the future evolution of the population. Among the discrete-time models, matrix models are often used to study the dynamics of structured populations (either age-structured or size-structured populations). They also permit to simplify the dynamics of a population into its basic components: recruitment or birth, growth or ageing, and mortality. The Usher model is a matrix model for size-structured populations that is based on four hypothesis (Usher, 1966, 1969):

- Hypothesis of independence: the evolution of the individuals are independent.
- Markov hypothesis: the evolution of an individual between two time steps t and $t+1$ depends only on its state at time t.

* Corresponding author.
E-mail addresses: melanie.zetlaoui@libertysurf.fr (M. Zetlaoui), nicolas.picard@cirad.fr (N. Picard), avner.bar-hen@smbh.univ-paris13.fr (A. Bar-Hen).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.07.031

- Usher hypothesis: during each time step, an individual can stay in the same class, move up a class, or die; each individual may give birth to a number of offspring.
- Hypothesis of stationarity: the evolution of individuals between two time step is independent of time.

A more general matrix model was proposed by Lefkovitch (1965), and allows any transition from one stage to another. The Leslie (1945) model, which describes a population grouped by age, is a special case of the Usher model: during each time step, an individual can only move up a class or die.

In this paper, we are interested in the Usher model and particularly with the Usher hypothesis. This hypothesis reduces the number of parameters of the model, compared with a more general matrix model such as the Lefkovitch model. When facing limited data, it is thus a useful alternative that permits more precise parameter estimates. The only restriction is that the dynamic has to match the constraints imposed by the hypothesis. This is generally the case for size-structured populations where growth can be described by a first-order process. On the contrary, this is rarely the case for stage-structured populations. The trick then is to choose classes and time step so that Usher hypothesis is realistic (Favrichon, 1998). These reasons can explain in particular why Usher models have been intensely used in forestry to model the dynamics of diameter distributions.

Nevertheless, even when using proper classes and time step, some data can violate Usher hypothesis. We consider here two ways of handling these data:

- Remove them prior to data analysis.
- Rectify them: individuals that move up by more than one class in a time step are considered as individuals that move up a single class, and individuals that move backwards are considered as individuals that stay in the same class.

We thus obtain two versions of transition parameters depending on whether data are trimmed or not.

The purpose of this study is to present a criterion of choice between these two data treatments, based on the robustness of the estimator of transition parameters. To compare the robustness of these two estimators of transition parameters, we use the influence curve. This gives a quantitative criterion of comparison of robustness of estimators . Furthermore, it gives the asymptotic behavior of an estimator. We compare then the asymptotic variance of the estimators and the difference between their expectations. Finally, the results are applied to a data set obtained from experimental forest plots in a tropical rain forest at Paracou (French Guiana).

2. Influence function

The approach to robust estimation of a location parameter by the influence curve was developed by Hampel (1974). It deals with estimators viewed as functionals. Let $(X_1, \ldots, X_n)$ be an n-sample with distribution function $F(\theta)$ on $\mathbf{R}^d$, with θ an unknown parameter. θ is supposed to express itself in the form $T(F)$. Then, the empirical estimator of θ is $T(F_n)$, where F_n is the empirical distribution function. The estimator is said to be generated by the functional T. Then, the influence curve, $IC_{T,F}(x)$, defined, under conditions of differentiability of T, by

$$IC_{T,F}(x) = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F + \varepsilon\delta_x) - T(F)}{\varepsilon},$$

where δ_x is a Dirac mass at x, describes the effect of an additional observation $x \in \mathbf{R}^d$ on the statistic $T(F)$. It gives the asymptotic behavior of the estimator

$$\sqrt{n}\,(T(F_n) - T(F)) \to \mathcal{N}\left(0, \int IC_{T,F}(x)^2 \, \mathrm{d}F(x)\right),$$

where the convergence is in law. Furthermore, the influence curve permits to obtain criteria to assess the robustness of an estimator. A first criterion is given by

$$\mu = \sup_x \left\| IC_{T,F}(x) \right\|_\infty,$$

where $\|.\|_\infty$ is the infinity norm. μ has been called the gross error sensitivity by Hampel (1974). It measures the biggest influence caused by a little perturbation on the value of the estimator.