

PARAMETER ESTIMATION FOR STAGE-DURATION MODELS

HOA THI THU PHAM

(Received 30 October 2018; first published online 7 January 2019)

2010 Mathematics subject classification: primary 62-07; secondary 62F30.

Keywords and phrases: survival analysis, multi-stage models, stage duration, stage frequency data, Laplace transform methods, destructive samples.

Multi-stage time-evolving models are fundamental for many biological systems in which time-evolving progression moves through distinct stages [1, 5, 7–9]. Generally, there are three classes of multi-stage models: matrix models, stage-duration models and delay-differential equation models [2]. This thesis focuses on stage-duration models for single-cohort stage-frequency data, which are usually applied to unmarked cohort data. In these models, individuals are assumed to enter the study population at the same time. In addition, individuals are not identified due to destructive sampling at different times. The data are collected by assessing the stages reached by individuals as a stage-structured time series. In particular, the numbers of alive and dead individuals in each stage are counted at each sampling time. These stage-duration data are fundamental to biology and have been studied in various biological contexts [4, 8–10]. Exploring the development of the population in each stage gives an understanding of maturation of individuals through their life cycle or industrial processes or the progression of a disease. Thus, these models are also important for developing strategies to treat diseases.

In stage-duration models, life cycles of individuals are divided into stages. The distributions of stage duration and survival time in each stage are commonly modelled using exponential, Weibull, Erlangian and gamma distributions. The main challenge of these models is to develop methods for estimating stage-dependent maturation parameters and hazard rates in each stage. Previous studies lacked flexibility in the assumptions needed to estimate the stage-dependent maturation parameters and the hazard rate parameters in each stage. When the stage durations were modelled using gamma distributions, the shape or the rate parameters of the distributions were

Thesis submitted to Flinders University in January 2017; degree approved on 29 May 2017; principal supervisor Jerzy Filar, co-supervisors Alan Branford, Murk Bottema and Darfiana Nur.

© 2019 Australian Mathematical Publishing Association Inc.

assumed to be known and the same. This assumption is frequently unrealistic, but it simplifies estimates. They also assumed that individuals could not die but rather would eventually move to the next stage, or that individuals could die but that the hazard rate was the same for each stage [2, 3, 6, 9].

In this thesis, the assumptions on the hazard rate parameters are relaxed and no assumptions are made on the maturation parameters. The objectives are to propose novel methods for estimating parameters in the following situations:

- to estimate parameters for models with stage-wise constant hazard rates by constant shape rate and constant scale cases;
- to estimate parameters for models with linear time-dependent hazard rates by constant shape rate and constant scale cases;
- to estimate parameters without initial knowledge of shape and rate parameters for models with no hazard rates;
- to estimate parameters without initial knowledge of shape and rate parameters for models with stage-wise constant hazard rates;
- to estimate parameters without initial knowledge of shape and rate parameters for models with linear time-dependent hazard rates.

This thesis is organised into six chapters including the introductory chapter. Chapter 2 introduces stage-duration models, the Laplace transform method, Bayesian analysis and previous studies of stage-duration models. In Chapter 3, we propose two new stage-duration models where nontrivial hazard rates apply. The first model considers hazard rates that are constant within each stage but vary between stages. The second model considers linear time-dependent hazard rates within stages. We use the Laplace transform method to estimate stage-dependent maturation parameters for these models. In addition, we propose methods for estimating these nontrivial hazard rates. A key result in this chapter is to explore the relationships among the stage-dependent maturation parameters in each stage. Those relationships are then embedded within the Markov chain Monte Carlo (MCMC) algorithm proposed in the subsequent chapter for better convergence of Markov processes. Each model is applied to simulated datasets in order to evaluate the accuracy. Chapter 4 uses a Bayesian approach to estimate parameters in the models with no hazard rates, with stage-wise constant hazard rates and with linear time-dependent hazard rates. The deterministic Metropolis–Hastings (MH) algorithm is used to improve the mixing of Markov chains. The main aim of this chapter is to relax assumptions about maturation parameters in the proposed models. By allowing uncertainties through prior distributions, parameters are estimated using a Bayesian approach. These methods are implemented in simulated datasets. In Chapters 5 and 6, we apply the techniques from Chapter 4 to simulation studies and the case studies of cattle parasite data and data related to breast development of New Zealand schoolgirls. Finally, Chapter 7 draws conclusions from the major results in the thesis and discusses remaining problems and future research directions.

References

- [1] H. Caswell, *Matrix Population Models: Construction, Analysis and Interpretation*, 2nd edn (Oxford University Press, Oxford, 2001).
- [2] P. De Valpine, K. Scranton, J. Knape, K. Ram and N. J. Mills, 'The importance of individual developmental variation in stage structured population models', *Ecol. Lett.* **17**(8) (2014), 1026–1038.
- [3] J. Hoeting, R. Tweedie and C. Olver, 'Transform estimation of parameters for stage-frequency data', *J. Amer. Statist. Assoc.* **98** (2003), 503–514.
- [4] C. Hopkins, G. Subramanian and H. Stallard, 'The development of *Hymenolepis diminuta* in primary and secondary infections in mice', *Parasitology* **64**(3) (1972), 401–412.
- [5] N. Keyfitz and H. Caswell, *Applied Mathematical Demography* (Springer, New York, 2005).
- [6] J. Knape and P. De Valpine, 'Monte Carlo estimation of stage structured development from cohort data', *Ecology* **97**(4) (2016), 992–1002.
- [7] B. Manly, *Stage-structured Populations: Sampling, Analysis and Simulation* (Chapman and Hall, New York, 1990).
- [8] K. Read and J. Ashford, 'A system of models for the life cycle of a biological organism', *Biometrika* **55**(1) (1968), 211–221.
- [9] H. Schuh and R. Tweedie, 'Parameter estimation using transform estimation in time-evolving models', *Math. Biosci.* **45**(1) (1979), 37–67.
- [10] R. Young, N. Anderson, D. Overend, R. Tweedie, K. Malafant and G. Preston, 'The effect of temperature on times to hatching of eggs of the nematode *Ostertagia circumcincta*', *Parasitology* **81**(3) (1980), 477–491.

HOA THI THU PHAM, An Giang University, 18 Ung Van Kiem Street,
Dong Xuyen Precinct, Long Xuyen City, An Giang Province, Vietnam
e-mail: ptthoa@agu.edu.vn