BOOK REVIEW

Parameter redundancy and identifiability, by Diana Cole

Frédéric Barraquand *

Most quantitative biologists and applied statisticians interested in *identifiability* — i.e., whether a unique set of parameter values can be found to maximize the likelihood of a model — currently have to swift through piles of manuscripts that are often only tangentially relevant to their needs. A colleague who fits dynamic models to data recently lamented: ‘Is there something that I could recommend to my graduate students and postdocs that would not be a very technical book from the 80s or 90s?’ Now there is. Diana Cole’s new book, *Parameter redundancy and identifiability*, explores all facets of model *identifiability* and its little cousin *parameter redundancy* (when the model is non-identifiable because it could be re-parameterized with fewer parameters). The book is written with mostly ecological modelling in mind (in the broadest sense — including fisheries science, epidemiology, ...). However, as the author notes in the preface, the methodology is general and applies equally to other fields: for instance, to compartmental models from pharmacokinetics, that are covered to some degree, and many other dynamical systems. In many ways, this book builds on the author’s previous work (Cole *et al.*, 2010; Cole & McCrea, 2016), but presents identifiability techniques in a more accessible and comprehensive manner than the original papers, which were more focused on so-called exhaustive summaries.

The book is richly illustrated with contrasted examples (binomial-based occupancy models, multinomial mark-recovery models, Gaussian state-space models for noisy count data, continuous-time compartmental models...), for which code is provided. Clear definitions are provided for concepts in the most theoretical chapters (chapters 3 to 6) — some definitions are more verbal than mathematical, but clear nonetheless. The definitions are then put to the test using case studies, including some fairly complex ecological models in chapter 9. I anticipate that the book will make for a wonderful graduate reading seminar series, going through one chapter at a time. I strongly recommend going through at least some of the case studies with pen, paper and code; I definitely found them very useful.

Despite these many virtues, on occasion I have found that the book was lacking mathematical precision. For instance, p. 44 it is stated that the derivative matrix $\mathbf{D}$, the Hessian matrix, and the Fisher Information matrix all have the same rank for exponential family models, with only a reference to the rather technical article of Catchpole & Morgan (1997). Given the fundamental nature of this result to justify the use of $\mathbf{D}$, a little more mathematics in Appendices would have helped the reader to bridge the gap between the book and the technical literature. Likewise, the Fisher Information Matrix for the basic occupancy model (a binomial model) p. 45 is provided only as Maple Code, but many readers may not use Maple, thus some analytical derivation would have been useful. Numerical computations are done in R which is great, perhaps a freely available symbolic tool as well (or more explicit mathematics) would be useful in next editions. To stay on the practical side, the book has many nice illustrations, though axes labels are sometimes a bit small.

Regarding literature coverage, the book is a treasure trove of up-to-date information on identifiability. I found myself avidly downloading cited papers, both theoretical and applied. In spite of that — somewhat paradoxically — there is little critical discussion of the validity of the concepts and their relation to the greater statistical literature. For instance, the notion of *practical identifiability* of Raue *et al.* (2009), which states that a parameter is identifiable when confidence regions do not escape to infinity, is introduced p. 38. This definition of practical identifiability somehow clashes with previous definitions that all refer to *point identifiability* (as in *point estimation*), not the existence of finite *intervals* (as in *interval estimation*, e.g. Casella & Berger 2002). Indeed, Raue *et al.* (2009)’s definition

*Frédéric Barraquand, Institute of Mathematics of Bordeaux, CNRS & University of Bordeaux, France (Email: frederic.barraquand@u-bordeaux.fr)

of unidentifiability does not preclude the existence of a unique point estimate and interacts with the shape of the likelihood far away from that point estimate. Other cited authors like Eisenberg & Hayashi (2014) have different definitions of practical identifiability that resemble instead Cole (2020)’s *extrinsic* identifiability, i.e., the existence of a unique point estimate for a model fitted to a given finite dataset. Therefore, although it does provide a comprehensive map of the identifiability literature, this book will not solve all of the literature’s ambiguities and contradictions, which were previously noted by Lavielle & Aarons (2016). Rather, the book’s emphasis is on popularizing technical papers and showing how to implement in practice the available methodology. A more internally consistent identifiability framework may eventually be desirable: the author favoured for now pedagogy over self-consistency.

I have appreciated the methodological pluralism of the book; pluses and minuses of the various available methods are clearly exposed. The only downside is that readers may at times hesitate regarding what to use and when – some information to help with this choice is provided in chapter 4 p. 97-100 and in chapter 10, but it is quite succinct. I would have liked a little more discussion. For example, in which case are exhaustive summaries most useful? They do seem very efficient for binomial and multinomial models, for which the dimensionality of $\mathbf{D}$ is low and defining summaries seems straightforward. For the dynamic models of chapter 8, it was more difficult to pinpoint how the author came to select the exhaustive summaries, and to what degree these are approximations (which they seem to be in several cases). That said, such imprecisions may simply reflect the state of an evolving research field.

These couple of limitations aside, the book is a well-written, self-contained and accessible introduction to the flourishing literature on identifiability; it will be an invaluable resource to all researchers interested in the topic for years to come. It is a very useful book both to better comprehend what is identifiability (either intrinsic, due to the model structure, or extrinsic, due to a particular data set) and to get a grip on the various techniques allowing to ascertain whether we can actually estimate the parameters of our models.

References

- Casella, G. & Berger, R.L. (2002). *Statistical inference*. 2nd edn. Duxbury Pacific Grove, CA.
- Catchpole, E.A. & Morgan, B.J. (1997). Detecting parameter redundancy. *Biometrika*, 84, 187–196.
- Cole, D. (2020). *Parameter Redundancy and Identifiability*. CRC Press.
- Cole, D.J. & McCrea, R.S. (2016). Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal*, 58, 1071–1090.
- Cole, D.J., Morgan, B.J. & Titterton, D. (2010). Determining the parametric structure of models. *Mathematical biosciences*, 228, 16–30.
- Eisenberg, M.C. & Hayashi, M.A. (2014). Determining identifiable parameter combinations using subset profiling. *Mathematical biosciences*, 256, 116–126.
- Lavielle, M. & Aarons, L. (2016). What do we mean by identifiability in mixed effects models? *Journal of pharmacokinetics and pharmacodynamics*, 43, 111–122.
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25, 1923–1929.