BOOK REVIEW:

RAY HILBORN AND MARC MANGEL'S THE ECOLOGICAL DETECTIVE: CONFRONTING MODELS WITH DATA

CARSTEN F. DORMANN

Monographs in Population Biology 28
Princeton University Press
Princeton, New Jersey
1997
ISBN: 0-691-03496-6 (hardcover) 0-691-03497-4 (paperback)
xvii + 315 pages
€27.65

The statistical ontogeny of an ordinary ecologist starts with regression and t-test and leads along ANOVA towards logistic regression, where in the far the peaks of "real statistics" become visible through the clouds. One of the reasons why those peaks too often stay far away is that we are seduced by clickable graphic user interfaces. Another one is that the mathematical representation alienates us from the underlying ecological question; and finally that we learn stats along the way, rather than "from scratch".

When I first held Hilburn & Mangel's Ecological Detective in my hands, in 1997, I was confused by the lack of "tests", at least things that would look like tests to me. Having only heard of, but not understood, maximum likelihood, I couldn't see what this book was about. Certainly all the illustrated problems could be solved by some "button clicking"? Now that I read rather than skimmed the book, my earlier puzzlement becomes explicable. It's an advanced statistics book, for those who already started climbing into the clouds of maximum likelihood, but haven't yet scaled "Mount Bayes".

The book's content is the illustration of how to use maximum likelihood estimations to answer ecological questions. As many ecological data deviate from normal distribution, the initially mentioned standard methods are not applicable: ordinary least-square regression and ANOVA are special cases of a bigger concept, maximum likelihood, for normally distributed underlying distributions.

The Ecological Detective gives a gentle introduction into Maximum Likelihood, which central line is:

$$\mathcal{L}(Y|\theta) = \prod_{i=1}^{n} f(y_i;\theta)$$

which is read as "the likelihood of the data Y, given the parameters θ of a model hypothesis, is the product over all data points of the probability distribution f of Y for possible values of θ". In common statistics f is the normal distribution, while for logistic regression it is the binomial distribution and so forth. Taking the logarithm of the above equation leads to the more commonly used log-likelihood function:

$$\mathbf{L}(Y|\theta) = log(\mathcal{L}(Y|\theta)) = \sum_{i=1}^{n} log(f(y_i;\theta))$$

It is this function that you have to maximise (mostly by using an iterative maximum searching algorithm, i.e. finding the maximum in a non-linear regression of the log-likelihood function on the data).

Moreover, the first seven chapters provide an overview of general statistical rules and advices. The take-home-message is KNOW YOUR DATA, which can only be achieved when the investigator understands the statistical background.

The chapters 6–9 step-by-step take the reader through model comparison and different likelihood functions (for normal, binomial, Poisson and lognormal error distributions). They illustrate the main procedures with a pseudocode to implement the method software-independent and exemplifies the main points with ecological case studies.

The final chapters (before a closing one on goodness of fit assessment and computational topics) are an introduction to Bayesian statistics. Hilburn & Mangel argue that all stats are in fact Bayesian, just with implicit definition of a "neutral" prior. You may recall that Bayes' theorem reads

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

which means that one can calculate the conditional probability of A on B according to the formula from the conditional probability of B on A and the probabilities of both A and B. Translated into Bayesian phrasing, the above formula can be read as

$$P(H_i|data) = \frac{\mathcal{L}(data|H_i)Prior(H_i)}{P(data)}$$

where $P(H_i|data)$ is the so-called posterior probability of the hypothesis, given the data. "The numerator is the joint probability of the data and H_i. The denominator must be the sum of such joint probabilities, summed over all possible hypotheses." (page 204). Thus the above can be re-written as

$$P(H_i|data) = \frac{\mathcal{L}(data|H_i)Prior(H_i)}{\sum_j \mathcal{L}(data|H_j)Prior(H_j)}$$

where H_j are the hypotheses competing with H_i.

$Prior(H_i)$ allows the incorporation of additional information, such as expert knowledge, but more importantly previous data. A somewhat cynical example is the 1990 sentencing for murder of a suspect, solely based on a matched DNA. Knowing that the chances for a match are 1 : 3 000 000 sounds convincing. But if the DNA match was derived from a DNA database with 10 Mio. men's data (the prior, in which the chance that a single person is innocent is (10 Mio. − 1)/10 Mio.), we could re-calculate the probability of innocent given the DNA match as

$$P(innocent|DNAmatch) = \frac{1/3 \text{ Mio.} \cdot (10 \text{ Mio.} - 1)/10 \text{ Mio.}}{1/3 \text{ Mio.} \cdot (10 \text{ Mio.} - 1)/10 \text{ Mio.} + 1/10 \text{ Mio.}} = 0.77$$

With this number Bayesian statistics give a precise number to our intuition: in the database are 2 or 3 people with a matching DNA, thus the probability of any single one of them is the wanted murderer is only 0.77. Hilburn & Mangel again give an extended example in an ecological context.

Overall, the book has three messages (apart from the KNOW YOUR DATA and KNOW YOUR GOODNESS OF FIT): 1. Maximum Likelihood estimation is the method of choice for a wide variety of ecological statistical problems. 2. You need little programming skills to use Maximum Likelihood. And 3. If at all possible, familiarise yourself with Bayesian statistics, they bear great potential.

While all this is worth knowing as an ecologist, The Ecological Detectives makes no reference to the fact that in advanced stats software (such as SAS, S-plus/R) these Maximum Likelihood procedures are implemented and hence the given pseudocodes and techniques have a very purist touch.

Having said that, the book brilliantly illustrates how simple Maximum Likelihood Estimations can be, and how relevant it is to much of ecological statistics. Since I read The Ecological Detective I regularly write down the likelihood function on a piece of paper, before handing the computational burden over to a computer alright.

Carsten F. Dormann
Applied Landscape Ecology
UFZ Leipzig-Halle
Permoserstr.15
04318 Leipzig
Email: dormann@alok.ufz.de

Leipzig, March 4, 2003