A Computational Approach to Statistical Arguments in Ecology and Evolution

Scientists need statistics. Increasingly this is accomplished using computational approaches. Freeing readers from the constraints, mysterious formulas, and sophisticated mathematics of classical statistics, this book is ideal for researchers who want to take control of their own statistical arguments.

It demonstrates how to use spreadsheet macros to calculate the probability distribution predicted for any statistic by any hypothesis. This enables readers to use anything that can be calculated (or observed) from their data as a test statistic, and hypothesize any probabilistic mechanism that can generate data sets similar in structure to the one observed.

A wide range of natural examples drawn from ecology, evolution, anthropology, paleontology, and related fields give valuable insights into the application of the described techniques, while complete example macros and useful procedures demonstrate the methods in action and provide starting points for readers to use or modify in their own research.

George F. Estabrook is a Professor of Botany in the Department of Ecology and Evolutionary Biology at the University of Michigan, Ann Arbor. He is interested in the application of mathematics and computing to biology, and has taught graduate courses on the subject for more than 30 years.

A Computational Approach to Statistical Arguments in Ecology and Evolution

GEORGE F. ESTABROOK

University of Michigan, Ann Arbor

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107004306

First published 2011

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-00430-6 Hardback

Additional resources for this publication at www.cambridge.org/9781107004306

Contents

Acknowledgments

In the mid 1980s, Neal Oden quit his boring job as a computer programmer and joined the Ph.D. program of the Department of Biology at the University of Michigan to become my eighth Ph.D. student. Neal had read a journal article in which the authors claimed that when data sampled a given probability distribution, the predicted distribution of a statistic they had invented would asymptotically converge to a known pre-calculated distribution of classical statistics. From the authors' explanation, it was not clear to Neal why. So he wrote a computer program that used a random number generator to sample the given distribution to create a large data set, and calculate the statistic; his program repeated this 1000 times to estimate the predicted distribution. The middle 80% was pretty close to the classical distribution, but upper and lower 10% differed. By doing this, Neal showed me how easy it is to make accurate estimations of the probability distribution for any statistic predicted by any hypothesis of random. Thank you, Neal. I began to apply Neal's idea in collaboration with other colleagues, and eventually offered a course to graduate students.

The many students who took my course, BIO 480 and later EEB 480, have contributed much to the development of this book, which has gone through a half dozen unpublished editions over the past decade in response to their helpful suggestions and requests.

One of the major stumbling blocks for students trying to use these concepts and techniques has been the need to learn enough computer programming to implement them. I taught my course originally using PASCAL, a programming language designed explicitly to teach students good programming practices. Even with PASCAL, the INPUT, OUTPUT, and MEMORY MANAGEMENT details were especially challenging for students who had never programmed before. A few years ago, my 18th Ph.D. student, Don Schoolmaster, showed me how to program spreadsheet macros that read from, and write to,

spreadsheets. This eliminates problems with INPUT, OUTPUT, and MEMORY MANAGEMENT, which makes the programming much more accessible to students. Virtually all students are familiar with spreadsheets, even if they have never programmed a macro. Thank you, Don.

Keith Hunley, a former EEB 480 student and now a professor at the University of New Mexico, read much of the pre-final draft of the book. He made hundreds of suggestions that improved the clarity and organization. Thank you, Keith.

My wife, Virginia Hutton Estabrook Ph.D., helped me with drawing the figures, organizing the literature cited, and grappling with the WINDOWS operating system. She also read some of the pre-final draft and made helpful suggestions. Thank you, Virginia.