Mathematical Statistics with Resampling and R

Laura Chihara
Carleton College

Tim Hesterberg
Google

Preface

Mathematical Statistics with Resampling and R is a one term undergraduate statistics textbook aimed at sophomores or juniors who have taken a course in probability (at the level of, for instance, Ross (2009), Ghahramani (2004) or Scheaffer and Young (2010)) but may not have had any previous exposure to statistics.

What sets this book apart from other mathematical statistics texts is the use of modern resampling techniques—permutation tests and bootstrapping. We begin with permutation tests and bootstrap methods before introducing classical inference methods. Resampling helps students understand the meaning of sampling distributions, sampling variability, P-values, hypothesis tests, and confidence intervals. We are inspired by the textbooks of Waldrop (1995) and Chance and Rossman (2005), two innovative introductory statistics books which also take a non-traditional approach in the sequencing of topics.

We believe the time is ripe for this book. Many faculty have learned resampling and simulation-based methods in graduate school and/or use them in their own work, and are eager to incorporate these ideas into a mathematical statistics course. Students and faculty today have access to computers that are powerful enough to perform resampling quickly.

A major topic of debate about the Mathematical Statistics course is how much theory to introduce. We want mathematically talented students to get excited about statistics, so we try to strike a balance between theory, computing and applications. We feel that it is important to demonstrate some rigor in developing some of the statistical ideas presented here, but that mathematical theory should not dominate the text. And of course, anytime additions are made to a syllabus, deletions must then also be made. Thus, some topics such as sufficiency, Fisher information and ANOVA have been omitted in order to make room for permutation testing, bootstrap and other modern computing methods (though we plan to make some of these omitted topics available as supplements on the text web page `https://sites.google.com/site/ChiharaHesterberg`).

We have compiled the definitions and theorems of the important probability distributions into an appendix (see Appendix B). Instructors who want to prove results on distributional theory can refer to that chapter. Instructors who wish to skip the theory can continue without interrupting the flow of the statistical discussion.

Incorporating resampling and bootstrapping methods requires that students use statistical software. We use R because it is freely available (`http://www.r-project.org/`), powerful, flexible, and a valuable tool in future careers. One of us works at Google where there is an explosion in the use of R, with more and more non-statisticians learning R (the statisticians already know it). We realize that the learning curve for R is high, but believe that the time invested in mastering R is worth the effort. We have written some basic materials on R that are available on the website for this text. We recommend that instructors work through the introductory worksheet with the students on the first or second day of the term, in a computer lab if possible. We also provide R script files with code found in the text, and additional examples.

Statistical computing is necessary in statistical practice and for people working with data in a wide variety of fields. There is an explosion of data—more and more data—and new computational methods are continuously being developed to handle this explosion. Statistics is an exciting field, dare we even say sexy?[1]

Acknowledgements This textbook could not have been completed without the assistance of many colleagues and students. In particular, we would like to thank Professor Katherine St. Clair of Carleton College who bravely class-tested an early rough draft in her *Introduction to Statistical Inference* class during Winter 2010. In addition, Professor Julie Legler of St. Olaf College adopted the manuscript in her *Statistical Theory* class for Fall 2010. Both instructors and students provided valuable feedback that improved the exposition and content of this textbook.

We would also like to thank Siyuan (Ernest) Liu and Chen (Daisy) Sun, two Carleton College students, for solving many of the exercises and writing up the solutions with LaTeX.

Finally, the staff at Wiley, including Steve Quigley and Jackie Palmieri, provided valuable assistance in preparing the manuscript for press.

[1] Try googling "statistics sexy profession."

Contents