Statistics 8701  
Computational Statistical Methods

Spring 2005

Gary W. Oehlert  
313B Ford  
625-1557  
gary@stat.umn.edu

Class hours: 9:05-9:55 MWF, Ford 115  
Office hours: 10:00-11:00 MWF, or by appointment

There will be no class Friday, Feb. 4 or the week of March 21.

Text: Givens and Hoeting *Computational Statistics*

Additional useful reference texts  
- Gentle *Elements of Computational Statistics*  
- Thisted *Elements of Statistical Computing*  
- Ripley *Stochastic Simulation*  
- Huber *Robust Statistical Procedures*  
- Efron *The Jackknife, the bootstrap, and other resampling plans*  
- Eubank *Spline Smoothing and Nonparametric Regression*

Required background: Statistics 8311 and computing experience. More about this below.

The book is new. It was originally supposed to be available in December, then January; it's currently in the "real soon now" category. I have preprints of a couple of chapters that I will distribute in a few days. I hope that the book will be here soon.

This course is computer intensive, as should be the case for a course in computational methods. We will use R and be writing our own programs in C or Fortran on the School of Statistics Linux workstations. Fifteen years ago I was a real whiz on S, but I have not kept up and do not have all the modern features of R at my fingertips (hey, with MacAnova, why do you need S?).

Speaking from experience, be prepared to spend some time to write your programs and much more time to debug your programs. This holds true for R and for languages such as C or Fortran. If you haven’t done any programming before, you will be spending a ton of time getting up to speed. Much, perhaps most, of our work can be done in R, but there will be a few problems for which R is simply not sufficiently efficient, and for those problems a C or Fortran program can save the day. Students have tried to do the computing in this course completely in a statistics package without programming; the results are generally disappointing.

**Grading** is based on assignments — there are no exams. Assignments will be due approximately one week after distribution. Assignments should be handed in as a regular paper assignments, but you can hand in code by giving a URL where it can be found on the web or a location in your home directory on the School of Statistics Linux network.
For the programming problems, you should hand in enough so that it is clear exactly what you
did and so that I can rerun your code if I wish.

If you get anything wrong on homework, you may rework the problem and hand in a fix.

Particularly when you are new to some programming environment, it is easy to get stuck (for
ever) because you don’t know some trivial little fact, flag, function, etc. In this class, it is OK to
ask for help! In this class, you may discuss the problems with me or with other students. In this
class, you may learn how to do a problem by watching someone else do the problem. But you can’t
turn in someone else’s work; it has to be work that you did, even if you had a lot of help to get to
the point that you could do it yourself.

It will generally be easier to do the homework problems on a Unix machine (our departmental
Linux boxes, or Mac OS X). If you want to do them some other, that is to say, harder way (for
example, you have a Windows box at home), you may do so, but I may not be able to help you.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming and R</td>
<td>R documentation</td>
</tr>
<tr>
<td>Numerical analysis preliminaries</td>
<td></td>
</tr>
<tr>
<td>Floating point numbers</td>
<td>Thisted, Ch. 2</td>
</tr>
<tr>
<td>Numerical integration</td>
<td>G&amp;H, Ch. 5; Thisted, Ch. 5</td>
</tr>
<tr>
<td>Optimization</td>
<td></td>
</tr>
<tr>
<td>Standard methods</td>
<td>G&amp;H, Ch. 2; Gentle, Ch. 1</td>
</tr>
<tr>
<td>Combinatorial optimization</td>
<td>G&amp;H, Ch. 3; Thisted, Ch. 4</td>
</tr>
<tr>
<td>EM Algorithm</td>
<td>G&amp;H, Ch. 4</td>
</tr>
<tr>
<td>Generating pseudorandom deviates</td>
<td></td>
</tr>
<tr>
<td>Uniforms</td>
<td>Gentle, Ch. 2.1; Ripley, Ch. 2</td>
</tr>
<tr>
<td>Other distributions</td>
<td>G&amp;H, Ch. 6; Gentle, Ch. 2.1; Ripley, Ch. 3</td>
</tr>
<tr>
<td>Classical Monte Carlo Methods</td>
<td>G&amp;H, Ch. 6; Gentle, Ch. 2; Ripley, Ch. 5</td>
</tr>
<tr>
<td>Markov Chain Monte Carlo</td>
<td>Geyer’s notes</td>
</tr>
<tr>
<td>Preliminaries</td>
<td>G&amp;H, Ch. 7; Gentle, Ch. 2; Geyer’s notes</td>
</tr>
<tr>
<td>Markov chain Monte Carlo</td>
<td>G&amp;H, Ch. 8</td>
</tr>
<tr>
<td>More topics</td>
<td></td>
</tr>
<tr>
<td>Direct assessment and resampling</td>
<td></td>
</tr>
<tr>
<td>Jackknife</td>
<td>Gentle, Ch. 3.3; Efron, Ch. 2—4</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>G&amp;H, Ch. 9; Gentle, Ch. 4; Efron, Ch. 5, 7</td>
</tr>
<tr>
<td>Cross validation</td>
<td>Gentle, Ch. 3.2; Efron, Ch. 7</td>
</tr>
<tr>
<td>Robust estimation</td>
<td>Gentle, Ch. 11.3; Huber, Ch. 1—3</td>
</tr>
<tr>
<td>Nonparametric regression</td>
<td></td>
</tr>
<tr>
<td>One predictor</td>
<td>G&amp;H, Ch. 11; Gentle, Ch. 6; Eubank, Ch. 2–5</td>
</tr>
<tr>
<td>Series, kernels, splines, wavelets</td>
<td></td>
</tr>
<tr>
<td>Several predictors</td>
<td>G&amp;H Ch. 12; Gentle, Ch. 10–11</td>
</tr>
<tr>
<td>PPR, ACE, trees, Mars, etc.</td>
<td></td>
</tr>
<tr>
<td>Bagging, boosting, bumping</td>
<td>notes</td>
</tr>
<tr>
<td>Density estimation</td>
<td>G&amp;H, Ch. 10; Gentle, Ch. 9</td>
</tr>
</tbody>
</table>

This material is available in alternative formats upon request. Please contact Mary or Dana, School
of Statistics, 313 Ford, 57300.