This course focuses on the modern statistical analysis of discrete choices of the form, “There are K options and you must choose one.” Inevitably, an empirical model of such discrete choices revolves around the computation of Probability[ε₁>max{ε₂,…,εₖ}]. In first-generation choice models, very special probability distributions were forced on us to make the multi-dimensional integration of this probability computational feasible (as in the multinomial logit model). With the recent advent of Markov chain Monte Carlo simulation procedures, the new generation of choice models are not restricted to such special distributions. Not only can we estimate the maximum likelihood of more satisfying models, but the simulation capability has dramatically increased the applicability of Bayesian estimation. This course will develop skills in applying modern modeling tools to real empirical data of discrete choices, both through homework exercises and a term project. In addition, the student will gain experience using the statistical programming language R, which has become a de facto standard among statisticians for the development of statistical software (R is the free version of the commercial S+ language).

Prerequisite: A doctoral level statistics or econometrics course.

Textbook: Train, Kenneth (2009), Discrete Choice Methods with Simulation, Cambridge University Press. This book may be bought from online sellers such as Amazon, paperback about $33, but Train makes pdf versions of the chapters available for free on his webpage, http://www.econ.berkeley.edu/~train/distant.html. At the same location are his videotaped lectures based upon the book.

Additional Readings:

Homeworks: To develop expertise in empirical analysis of choice data, you must experience data analysis. This will also allow you to become proficient with R. I will therefore ask you to do multiple homework assignments. The problem sets can sometimes be very challenging. To promote efficient learning, follow “Hess’s Rule”: if you have not been able to answer a problem after working on it for three hours, you must call me - at the office, at home, in the morning, or late at night - to get a hint or two. Seriously! My office and home phone numbers are listed above, but I am also in the telephone directory.

How good are your skills at typing mathematical and statistical equations? To enhance them, I’d like you to write-up your homework exercises answers using MS Word, using the Equation Editor to express mathematical formulas. There is a nice tutorial with videos at http://www.ist.uwaterloo.ca/ec/equations/equation.html. You can cut and paste output from R into Word. Submit hardcopy in class, not electronically online or via Blackboard Vista.
R Software: The statistical software language R is freely available at the Comprehensive R Archive Network (CRAN) http://www.cran.r-project.org. You should download and install R on your own computer. In session 2, we will begin to use R.

Computer Labs: Almost every class meeting I will be doing data analysis with R to repeatedly demonstrate that the abstract analysis of the textbook becomes real when data is confronted. In addition, several class sessions will be devoted to computer analysis. Please bring your laptop computer to class on those days so that you can participate in the computer lab.

Term Project: There are several hundred datasets on choice models available in R and textbooks; a list is included in our Blackboard bulletin board. You should select one of these that interests you and develop R software that implements a modern choice model. There is no rush to decide on the dataset and topic; on March 21, please turn in a one page statement of your research plans. I would like to meet with you later that week to discuss your project. You will get to communicate your analysis to the class as you would at an academic conference (15 minute presentation followed by 5 minutes Q&A) and in a paper (10 page maximum, double-spaced, excluding figures and tables).

Course Grading: Like in all doctoral courses, our goal is for you to do research that is publishable in the top academic journals (whose editorial boards are also receiving submissions from the best “veteran” researchers in the world). If your statistical analysis looks inexpert, how will you convince the editor that you have truly uncovered surprising but valid linkages between variables? I really want you to succeed in this endeavor, so I will provide evaluations of problem sets and project from the perspective of a journal editor. Personal improvement is all that matters to you and to me, but because the university requires that I give you a grade from a standardized set of options, I will combine my evaluations using the following weights: homework problems 50%, project presentation 10%, and project paper 40%.

Blackboard Vista: We will use the Blackboard Vista as a bulletin board to facilitate electronic communication. On our MARK 8349, I will post datasets, SPSS simpleton’s guides, lecture notes for some topics and respond to your questions about statistical research. You can log onto Blackboard Vista from any computer that has Web access to http://uh.edu/webct/index.html.

Academic Honesty: The University of Houston Academic Honesty Policy is strictly enforced by the C. T. Bauer College of Business. No violations of this policy will be tolerated in this course. A discussion of the policy is included in the University of Houston Student Handbook, http://www.uh.edu/dos/hdbk/acad/achonpol.html. Students are expected to be familiar with this policy.

Accommodations for Students with Disabilities: The C. T. Bauer College of Business would like to help students who have disabilities achieve their highest potential. To this end, in order to receive academic accommodations, students must register with the Center for Students with Disabilities (CSD) (telephone 713-743-5400), and present approved accommodation documentation to their instructors in a timely manner.
<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 19</td>
<td>Introduction to Discrete Choice</td>
<td>Skim for now McFadden’s Nobel Prize Lecture in (2001) <em>AER</em></td>
</tr>
<tr>
<td>2</td>
<td>Jan 24</td>
<td><strong>Computer Lab</strong>: CRAN, R commander, data i/o, object manipulation, array and matrix operations, functions, if &amp; for, graphs, shaping choice data</td>
<td>Kuhnert and Venables, “An Introduction to R.” pp. 11-96, Croissant “Explanation of mlogit package”</td>
</tr>
<tr>
<td>3</td>
<td>Jan 26</td>
<td>Theory of Discrete Choice</td>
<td>Train 2</td>
</tr>
<tr>
<td>4</td>
<td>Jan 31</td>
<td>Probability and Statistics via R</td>
<td>Greene Appendix B</td>
</tr>
<tr>
<td>5</td>
<td>Feb 2</td>
<td>Regression Analysis via R</td>
<td>Long 2, Kennedy 1-3, Farnsworth, “Econometrics in R”</td>
</tr>
<tr>
<td>7</td>
<td>Feb 9</td>
<td><strong>Computer Lab</strong>: Programming in R, if &amp; for, vector calculations, random variables, optimization, generalized linear model</td>
<td>Kuhnert and Venables, pp. 97-107</td>
</tr>
<tr>
<td>9</td>
<td>Feb 16</td>
<td>Binary Logit models</td>
<td>Long 3</td>
</tr>
<tr>
<td>11</td>
<td>Feb 23</td>
<td>Simulation Based Statistics</td>
<td>Stern (1997) <em>JEL</em></td>
</tr>
<tr>
<td>12</td>
<td>Feb 28</td>
<td>Drawing from Densities</td>
<td>Train 9.1-9.3</td>
</tr>
<tr>
<td>13</td>
<td>Mch 2</td>
<td>Estimation of Multinomial Probit with simulation</td>
<td>Train 5.6</td>
</tr>
<tr>
<td>14</td>
<td>Mch 7</td>
<td>Estimation with simulation</td>
<td>Train 10</td>
</tr>
<tr>
<td>15</td>
<td>Mch 9</td>
<td><strong>Computer Lab</strong>: Multinomial Probit in R</td>
<td>Long 4,</td>
</tr>
<tr>
<td>17</td>
<td>Mch 23</td>
<td><strong>Computer Lab</strong>: Probit via MCMC Bayes</td>
<td>Imai and Dyk, MNP: R Package for Fitting the Multinomial Probit</td>
</tr>
<tr>
<td>18</td>
<td>Mch 28</td>
<td>Multinomial Logit models</td>
<td>Train 3</td>
</tr>
<tr>
<td>19</td>
<td>Mch 30</td>
<td>Mixed Logit</td>
<td>Train 6</td>
</tr>
<tr>
<td>20</td>
<td>Apr 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Apr 6</td>
<td>GEV/Nested Logit</td>
<td>Train 4</td>
</tr>
<tr>
<td>22</td>
<td>Apr 11</td>
<td>Marginal Effect on Probabilities, Moderation effects</td>
<td>Greene 23.4, Ai and Norton</td>
</tr>
<tr>
<td>23</td>
<td>Apr 13</td>
<td>Individual-level Parameters</td>
<td>Train 11</td>
</tr>
<tr>
<td>24</td>
<td>Apr 18</td>
<td>Endogeneity</td>
<td>Train 13</td>
</tr>
<tr>
<td>25</td>
<td>Apr 20</td>
<td><strong>Computer Lab</strong>: Bayesian choice modeling</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Apr 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Apr 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Apr 27</td>
<td>Student presentations</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>May 2</td>
<td>Student presentations</td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Readings:


K. Train, Qualitative Choice Analysis, Cambridge, MA, MIT Press, 1986, Ch. 8


C. Bhat, "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," working paper, Department of Civil Engineering, University of Texas, Austin, 1999, forthcoming in Transportation Research


