AREC 815: Experimental and Behavioral Economics

Problem Set 3

Department of Agricultural and Resource Economics
University of Maryland
Fall 2016

Problem Set 3 is due by the end of the day on November 3.

1. Let $u : \mathcal{R}_+ \to \mathcal{R}$ be a strictly increasing Bernoulli utility function. Show that:

   (a) $u(\cdot)$ exhibits constant relative risk aversion equal to $\rho \neq 1$ if and only if $u(x) = \beta x^{1-\rho} + \gamma$ where $\beta > 0$ and $\gamma \in \mathcal{R}$.

   (b) $u(\cdot)$ exhibits constant relative risk aversion equal to 1 if and only if $u(x) = \beta \ln x + \gamma$ where $\beta > 0$ and $\gamma \in \mathcal{R}$.

2. This remainder of the problem set focuses on estimating logit models of discrete choices in experiments. Before you begin, read Chapters 2 and 3 of Ken Train's textbook, *Discrete Choice Methods with Simulation*. The entire book is available for download on Train's website. All of the code for this problem set is based on code originally written by Professor Train for his graduate course.

   The data set `arec815ps3data.csv` contains (simulated) data on the decisions of 1000 subjects, each of whom answered five lottery choice questions similar to those used by Binswanger (1980). The same data are also contained in `arec815ps3data.dta`. The lottery choice questions are those included in the Kenya Life Panel Survey, and are described in Table 1.

   Table 1: Lottery Choice Decision Problems

	Option 1		**Option 2**		**Option 3**	
	Heads	Tails	Heads	Tails	Heads	Tails
Decision Problem 1	100	100	0	400	-	-
Decision Problem 2	30	340	100	100	0	400
Decision Problem 3	100	100	55	240	30	340
Decision Problem 4	30	230	60	170	90	110
Decision Problem 5	10	200	70	160	90	110

   In each lottery choice question, a subject indicated which of two or three lotteries she preferred to play for real money. Each lottery represented the payouts associated with a flip of a coin: one payout associated with heads, and one payout associated with tails.

   Open the data set in Stata. As you can see, the data set contains 14,000 observations. Each observation is a lottery j in a choice situation s facing an individual i. In other words, the rows of the data set are the options that subject i might have chosen in some decision problem. The data set contains the following variables:

- **id**: ID number for the subject (from 1 to 1000)
  choicenum: ID number of the lottery choice problem as represented in Table 1 (from 1 to 5)
  sitid: ID number for the "choice situation" in which subject i faced choice problem s (from 1 to 5000)
  option: ID number for the option (i:e: the lottery — as represented in Table 1) within in choice situation (from 1 to 3)
  chosen: a dummy for the lottery chosen by i in a given choice situation
  heads: the heads payoff in a given lottery
  tails: the tails payoff in a given lottery
  exvalue: the expected value of a given lottery
  optionA: a dummy indicating whether a lottery was presented first (as in: at the top of the list) in the choice situation
  norisk: an indicator for degenerate lotteries yielding payoff y_{sj} with certainty

Explore the data set. What fraction of subjects chose the degenerate lottery that paid 100 dollars with probability one over the lottery that paid 400 dollars with probability 0.5 in Choice 1?

3. The matlab program `logit_q3.m` estimates a logit model of individual choices where the probability that subject i chooses lottery j in choice situation s is given by

$$P_{isj} = \frac{e^{\beta' x_{isj}}}{\sum_{k \in J} e^{\beta' x_{isk}}} \tag{1}$$

and x contains three variables: `exvalue`, `optionA`, `norisk`. In order for the code to run properly, the first column of `arec815ps3data.csv` must contain the `sitid` variable, and the data must be sorted by `sitid`; the second column must contain the `option` variable, and the data must be sorted accordingly within each choice situation; and the third column must contain the `chosen` variable. The other columns can contain any data you might want to use in your estimation.

First, read through both `logit_q3.m` and `loglike_q3.m`. Make sure you understand what is being done at each step of the estimation process. Then, run `logit_q2.m` to estimate the logit model. Report your `coefficient` estimates and standard errors.

4. You can also estimate simple logit models in Stata. Use the `asclogit` command to estimate the same simple logit model, in which `exvalue`, `optionA`, `norisk` are included as explanatory variables. You'll need to use the options `case(sitid)`, `alternative(option)`, and `noconstant`. Compare your results to those reported in Question 3. How does Stata's run time compare with that of matlab?

5. In practice, we think that choices between risky prospects are driven by risk aversion. Suppose subjects' preferences over risky outcomes can be represented by a utility function of the constant relative risk aversion form:

$$u(x) = \frac{x^{1-\rho}}{1-\rho}. \tag{2}$$

If subjects are homogenous (i.e. characterized by a single ρ parameter), then the probability that i chooses lottery j is given by

$$P_{isj} = \frac{e^{EU_{isj}/\sigma}}{\sum_{k\in J} e^{EU_{isj}/\sigma}} \tag{3}$$

where

$$EU_{isj} = \frac{1}{1-\rho}\left[\frac{1}{2}\left(y_{sj}^{heads}\right)^{1-\rho} + \frac{1}{2}\left(y_{sj}^{tails}\right)^{1-\rho}\right] \tag{4}$$

and σ is a a noise parameter which characterizes the relative importance of the deterministic and random components of the utility function. (In the model involving a linear $\beta' x$, β and σ are not separately identified, so the elements of β characterize the importance of factors as compared with the random component of utility.)

Revise the matlab code to estimate the parameters ρ and σ via maximum likelihood. Specifically, you will need to do three things. First, modify the section of `logit_q2.m` that identifies the independent variables of interest and names the parameters:

```
IDV=[6 7 8]; % Independent variables
NAMES={'expected value' 'option A' 'norisk'}; % Labels for independent variables
B=[0.5 0.5 0.5]; % Starting values for coefficients
```

You are now interested in using the payoffs in the heads and tails states as independent variables and calculating ρ and σ. Second, change the name of the likelihood function program in the line that calls the `fminunc` command in matlab. Third, when you revise `loglike_q2.m` to create a new program `loglike_q5.m`, you'll need to replace the line

```
v=VARS*coef;
```

and replace it with code that calculates the CRRA expected utility for each lottery and then divides the CRRA EU by the noise parameter, σ. You do not have to change the part of the likelihood that uses those inputs to calculate the logit probabilities or the overall log likelihood.

When you think you've modified the matlab programs successfully (i.e. you have them running), use the Stata program `gensimdata_q5.do` to generate a simulated data set with known ρ and σ parameters. Use your simulated data set to confirm that your matlab code recovers (approximately) correct model parameters. Iterate as needed, and describe your results and process.

6. Use your matlab program to estimate the CRRA coefficient underlying the choices made by subjects in `arec815ps2data.csv`. Report your results. What do the coefficient estimates suggest about the choices made by subjects?

7. Modify the program `gensimdata_q5.do` to simulate the choices of subjects with the ρ and σ parameters that you estimated in Question 6. Which lotteries are preferred by your simulated subject in expected (deterministic) utility terms? (One way to arrive at the answer quickly is to rerun your simulation code with a substantially lower noise parameter.)

8. In your simulated data generated using the ML estimates of ρ and σ, deviations from these deterministically-preferred choices should be driven by the additive random component of utility, which is uncorrelated across a given individual's decisions in the model. Explore this issue by generating two individual-level variables that indicate whether an individual subject chose the most risky option in each of the first two choice situations. In both the original data and your first simulated data set, explore the relationship between these two variables by means of regression or a t-test. Be sure that you are doing this at the individual and not the option level (your regressions should each have one thousand observations). Discuss your findings. What does this suggest about our exercise in maximum likelihood estimation?