Stanford PS350c, Jonathan Wand, Spring 2006

Problem set 1: due in class at beginning of lecture, April 24

Problemset structure. Make sure to hand in

- Your written results, clearly identifying the question number, in order
- Your figures, clearly identifying the question number
- Your **R** code, separately

Notes on **R**:

- I don't want you to get bogged down in **R**, so there are hints here, within the questions, and on the course website (*including* stubs of code for you to fill in/build on.)
- Useful functions for densities/random numbers: dnorm/rnorm, dunif/runif, dlogis/rlogis, dexp/rexp.
- When plotting density functions, consider defining support of data, and then feed this into plotting function. Eg, $x < -seq(-5, 5, .1); plot(x, rnorm(x)); lines(x, rnorm(x, 1, 2))$.

Fearon and Laitin replication:

- We will consider replication data from the article,

  James D. Fearon and David D. Laitin, "Ethnicity, Insurgency, and Civil War," American Political Science Review 97, 1 (March 2003): 75-90.

- I will simply refer to this as F/L. There is a link to the article and a link to R code, `http://www.stanford.edu/class/polisci350c/R/fearon_laitin/rerun_fl.R`
- NOTE: if you maintain a net connection while working on this material, you just need to use the code in the "just the R code link" which you can paste into an R session or an editor.
- The R code reproduces the results in their Table 1, column 3.

Notes on likelihood exercises

- I have posted stubs of code for these exercises at `http://www.stanford.edu/class/polisci350c/R/logit`
- These files will help you to focus on the conceptual ideas, rather than details of how to define functions.
- You will find that you can re-use this same framework in the future, so make sure to annotate your code!

Part I: Important Distributions, Analytics and Simulations

1. Based on Yellot and McFadden

   Let x_1 and x_2 be fixed scalar values and define the pdf, $\lambda(x) = e^{-x}e^{-e^{-x}}$, $-\infty \leq x \leq \infty$.

   Let $\epsilon_1 \sim \lambda(x) \quad \text{and} \quad \epsilon_2 \sim \lambda(x)$

   (a) Derive the cdf, $P(x_1 + \epsilon_1 < x)$

   (b) Derive the cdf, $P(x_2 + \epsilon_2 < x_1 + \epsilon_1)$

   Hint: you can define $\Delta = x_1 - x_2$, and construct the cdf based on

   $$\int_{-\infty}^{\infty} \lambda(\epsilon_1) \left[\int_{-\infty}^{\epsilon_1+\Delta} \lambda(\epsilon_2)\partial\epsilon_2 \right] \partial\epsilon_1$$

   but explain why this is true.

   (c) What is the pdf of the cdf in (b)? Ie,. differentiate the cdf in (b) with respect to Δ.

   (d) Provide (overlayed) plots of the pdf and cdf in (b) and (c).

2. You will be asked to simulate draws from $\lambda(x)$. How will you do it since **R** only has basic distributions coded up as functions?

   You can always simulate a distribution if you know the cdf and have a random generator for the uniform. Let the cdf be $F(x)$. The cdf maps x onto the unit interval, and the inverse of this function maps from the unit interval back into x.

   Example, consider the exponential distribution, $p = F(x) = 1 - e^{-\alpha x}$. The inverse function is $x = F^{\leftarrow}(p) = -\log(1-p)/\alpha$. Evaluating the inverse function using p from a uniform(0,1) random number generator will produce draws from the distribution of x.

   (a) Verify this procedure for the exponential; for example, plot density of simulated values against **R** standard function dexp().

   (b) Derive the inverse function for the cdf of the Gumbel distribution, and plot via simulation (as in (a))

Part II: Applications of Logit, Probit, Linear Probability

1. Relationships between logit and probit parameters.

   Thinking of models only as approximations, a researcher may adopt a quasi-likelihood interpretation of the estimation results. If so, one may pose the question of how parameter estimates of logits and probits relate. Using the F/L data and model in Table 1 column 3:

   (a) How do the parameters differ between logit and probit? Does your inference differ?

(b) Do you see any warnings in the R output? Can you explain why logit and probit might differ in this regard?

2. Some people would consider running a linear probability model for F/L data. Diagnose the problems that exist in doing so for the previous simulation, and analytically explain them.

3. (a) summarize assumptions necessary to construct Figure 2 in F/L

   (b) comment on assumptions, and note any concerns

   (c) replicate Figure 2

   (d) compare with a simple and average finite difference case study

4. For the log-likelihood of the logit model, $\Lambda(\alpha + \beta x)$,

   (a) derive the analytical first derivatives with respect to the vector $[\alpha, \beta]$

   (b) derive the 2x2 matrix of second derivatives with respect to $[\alpha, \beta]$.

5. Moral of the following story: need to know enough to know when the computer is doing you wrong.

   Consider the following table

	y=0	y=1
x= 0	10	10
x= 1	0	10

   which could be recreated with,

```
x <- c( rep(0,20), rep(1,10)); y <- c( rep(0,10), rep(1,20))
table(x,y)
```

   (a) Run LS, logit, probit with the model $F(\alpha + \beta x)$. Anything odd about these results, do they lead you to the same inference?

   (b) Analytically consider the first order conditions for maxizing the logit. What do the FOC with respect to β imply for the probability $P(y = 1 \mid x = 1)$? How would this value be approximated by the parameters in this logit parameterization?

   (c) Given (b), consider/comment on the 2x2 Hessian of this logit given the data. What implications does this have for any attempt at standard error calcuations of β in the logit for this data. Note: this is a simple enough case to actually work out the values manually.

   (d) Explain what is going on with the computer. Do you make the correct inference?

Part III: Likelihood

The goal of this problem set is to provide you with applications of the following concepts we have covered analytically, including (a) gradients and Hessian of log-likelihood; (b) information matrix (in)equality under (mis)specification, and MLE SE; and (c) invariance and bias of MLE

Except where noted, we will focus on a simple logit model for dichotomous choice,

$$P(y_i = 1) = \frac{1}{1 + \exp\{-\beta_0 + \beta_1 x_i\}}$$

where β_0, β_1, and x_i are all scalars, and $\beta = (\beta_0, \beta_1)$.

We will work with the **negative** of the log-likelihood (which I may refer to as -LL). Thus the estimation is going to be changed from a maximization problem to a minimization problem.

1. (optional **warmup.R**) The **R** code sketches rely on functions and on vector manipulations and scoping rules. If you are not familiar with this, I have posted a set of code you may wish to walk through slowly to build some intution.

2. Estimation. Up to now you have probably used `glm( ... ,family=distribution)` to estimate the parameters of a model, where `distribution` selects a likelihood/density for the dependent variable, such as binomial, or gaussian. If the likelihood is not one that is defined in the `glm(...,family=?)`, we can solve the non-linear optimization problem of finding the maximum likelihood estimates of a vector of paramers $\hat{\theta}$ using `optim()`. In **R** use `help(optim)` to get more details of this function.

   On the course site, I provide a skeleton of the code which provides an outline and starting points for the steps you will need to complete in order to go about this; you will need to fill in the missing parts.

   The data is on the website, and loaded in the sketch of code.

   (a) define the negative log-likelihood (-LL) function

   (b) define the score function of -LL (so keep track of the negative!)

   (c) define the hessian (so keep track of the negative!)

   (d) estimate the parameters using `optim()`,

   (e) calculate the SE using

      i. numerical Hessian (this will be of -LL from optim(), see code)

      ii. empirical Hessian (Long, p58, Var_2)

      iii. BHHH / outer-product-of-the-gradients (OPG) (Long, p58, Var_3)

      iv. (optional) the Huber-White / sandwich covariance matrix

   Describe/comment on any differences.

3. Profile the likelihood and score functions

   Fix the value of β_0 at the MLE. Make three plots varying the value of β_1 from -2 to 2 (the x-axis), with the y-axes for the values of each of the following

   (a) negative log-likelihood

   (b) score of β_0 of -LL

   (c) score of β_1 of -LL

   Comment on what you find.