Problem Set 2 Solutions

Tyler Ransom

Directions: Answer all questions. Clearly label all answers. Show all of your code. Compute standard errors for all parameter estimates. Turn in the following to me via your dropbox (in a folder labeled 'MatlabPS2.2') in Sakai by 11:59 p.m. on Thursday, August 2, 2012:

- m-file(s)
- .sh shell script file that executes your m-file(s)
- a log file (from off the cluster)
- matsub.oXXXXX file
- pdf version of your writeup with its LaTeXsource code

Put the names of all group members at the top of your writeup (each student must turn in his/her own materials).

1. Gradient of Multinomial Logit.

   (a) Following Train (p. 63), but adapting his derivation for the partial with respect to β_j instead of the whole vector β:

$$
\begin{aligned}
\frac{\partial \ell}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j}\sum_i\sum_j\sum_t d_{ijt}\ln\left(\frac{\exp\left(X_{ijt}\beta_j\right)}{\sum_k \exp\left(X_{ikt}\beta_k\right)}\right) \\
&= \frac{\partial \sum_i\sum_j\sum_t d_{ijt}\left(X_{ijt}\beta_j\right)}{\partial \beta_j} - \frac{\partial \sum_i\sum_j\sum_t d_{ijt}\ln\left(\sum_k \exp\left(X_{ikt}\beta_k\right)\right)}{\partial \beta_j} \\
&= \sum_i\sum_t d_{ijt}X_{ijt} - \frac{\partial}{\partial \beta_j}\sum_i\sum_j\sum_t d_{ijt}\ln\left(\sum_k \exp\left(X_{ikt}\beta_k\right)\right) \\
&= \sum_i\sum_t d_{ijt}X_{ijt} - \frac{\partial}{\partial \theta_j}\sum_i\sum_t \ln\left(\sum_k \exp\left(X_{ikt}\theta_k\right)\right)\left[\sum_j d_{ijt}\right] \\
&= \sum_i\sum_t d_{ijt}X_{ijt} - \sum_i\sum_t \frac{\partial}{\partial \beta_j}\ln\left(\sum_k \exp\left(X_{ikt}\beta_k\right)\right)[1] \\
&= \sum_i\sum_t d_{ijt}X_{ijt} - \sum_i\sum_t \frac{1}{\sum_k \exp\left(X_{ikt}\beta_k\right)}\frac{\partial}{\partial \beta_j}\left(\sum_k \exp\left(X_{ikt}\beta_k\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_i \sum_t d_{ijt} X_{ijt} - \sum_i \sum_t \frac{X_{ijt} \exp\left(X_{ijt}\beta_j\right)}{\sum_k \exp\left(X_{ikt}\beta_k\right)} \\
&= \sum_i \sum_t d_{ijt} X_{ijt} - \sum_i \sum_t P_{ijt} X_{ikt} \\
&= \sum_i \sum_t \left(d_{ijt} - P_{ijt}\right) X_{ijt}
\end{aligned}
$$

In matrix form, this is

$$\frac{\partial \ell}{\partial \beta_j} = X_j' \left(d_j - P_j\right)$$

where X_j is a matrix of dimension $(N*T) \times K_j$, β_j is a column vector of dimension $K_j \times 1$, and d_j and P_j are column vectors of dimension $(N*T) \times 1$.

2. Estimation of Multinomial Logit with and without gradient.

   (a) See Table 1.

   (b) See Table 2.

   (c) See m-file code and Table 3. Results are very similar with and without user-supplied gradient. However, starting from the same values, the gradient version took 29 seconds, whereas the non-gradient version took over 8 minutes.

3. Constrained optimization within `fminunc`.

   (a) See m-file code and Table 4.

4. Estimation of normal MLE with and without gradient and hessian.

   (a) The objective function I chose to work with is $\ell = -(n/2)\ln\left(\sigma^2\right) - \left(1/2\sigma^2\right)(y - X\beta)'(y - X\beta)$. If you transformed the objective function aside from taking the log, this will affect the hessian, and hence your standard errors. I chose to estimate σ rather than σ^2. Hence, for my problem, the (negative) gradient vector is slightly different from what is reported in Hayashi or Greene:

   $$-\begin{bmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \frac{-X'(y-X\beta)}{\sigma^2} \\ \frac{n}{\sigma} - \frac{(y-X\beta)'(y-X\beta)}{\sigma^3} \end{bmatrix}.$$

   See my m-file code, as well as Table 5, for how this performed.

   (b) Following part (a), the hessian for my problem is

   $$\begin{bmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell}{\partial \beta \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \beta'} & \frac{\partial^2 \ell}{\partial (\sigma)^2} \end{bmatrix} = \begin{bmatrix} \frac{X'X}{\sigma^2} & \frac{4X'(y-X\beta)}{\sigma^3} \\ \frac{4(y-X\beta)'X}{\sigma^3} & -\frac{2n}{\sigma^2} + \frac{6(y-X\beta)'(y-X\beta)}{\sigma^4} \end{bmatrix}.$$

   See my m-file code, as well as Table 6, for how this performed. The results are very similar regardless of how much analytical detail I provided `fminunc`. The main difference was convergence time. The hessian converged fastest, followed by gradient, followed by neither. Looking at 2(c), it is clear that once the model becomes more non-linear and/or increases in parameters, there are significant computational gains to be had from providing `fminunc` with more analytical information.

Table 1: Tabulation of Choice Set

Choice	Frequency	Percentage
School only	2,833	8.75%
Work in school	6,591	20.36%
Work PT (no school)	4,520	13.96%
Work FT (no school)	11,584	35.79%
Military	846	2.61%
Other	5,996	18.52%

Table 2: Multinomial Logit Estimation—No User-Supplied Gradient

(a) Multinomial Logit Point Estimates

	(school)	(schoolWrk)	(workPT)	(workFT)	(military)
constant	-2.071	-2.110	-0.741	-1.368	-0.764
male	0.373	-0.096	-0.105	0.453	1.524
AFQT	0.518	0.570	0.143	0.217	0.734
Mhgc	0.067	0.015	0.021	0.000	0.005
hgc	0.027	0.116	-0.107	-0.044	-0.474
exper	-0.943	-0.827	0.281	0.437	0.237
Diploma	1.571	2.035	1.021	1.225	3.617
AA	0.441	0.703	1.151	1.290	0.783
BA	-0.514	-0.528	1.277	1.922	1.000
log likelihood			-33,947.3		
iterations			386		

(b) Multinomial Logit Standard Errors

	school SE	schoolWrk SE	work PT SE	work FT SE	military SE
constant	0.268	0.227	0.210	0.177	0.479
male	0.052	0.044	0.041	0.035	0.092
AFQT	0.032	0.026	0.026	0.022	0.046
Mhgc	0.010	0.008	0.008	0.007	0.015
hgc	0.019	0.016	0.017	0.014	0.040
exper	0.025	0.019	0.012	0.011	0.024
Diploma	0.101	0.089	0.061	0.051	0.183
AA	0.147	0.124	0.133	0.119	0.260
BA	0.118	0.099	0.109	0.093	0.244

Table 3: Multinomial Logit Estimation—User-Supplied Gradient

(a) Multinomial Logit Point Estimates

	(school)	(schoolWrk)	(workPT)	(workFT)	(military)
constant	-2.030	-2.071	-0.710	-1.338	-0.717
male	0.372	-0.097	-0.105	0.452	1.523
AFQT	0.520	0.571	0.145	0.219	0.737
Mhgc	0.067	0.015	0.021	0.000	0.005
hgc	0.024	0.113	-0.109	-0.046	-0.477
exper	-0.945	-0.828	0.281	0.436	0.236
Diploma	1.577	2.039	1.024	1.229	3.618
AA	0.444	0.705	1.152	1.292	0.785
BA	-0.508	-0.523	1.282	1.927	1.007
log likelihood			-33,947.3		
iterations			24		

(b) Multinomial Logit Standard Errors

	school SE	schoolWrk SE	work PT SE	work FT SE	military SE
constant	0.268	0.227	0.209	0.177	0.478
male	0.052	0.044	0.041	0.035	0.092
AFQT	0.032	0.026	0.026	0.022	0.046
Mhgc	0.010	0.008	0.008	0.007	0.016
hgc	0.020	0.016	0.017	0.014	0.040
exper	0.025	0.019	0.012	0.011	0.024
Diploma	0.101	0.089	0.061	0.051	0.182
AA	0.147	0.124	0.133	0.119	0.260
BA	0.118	0.099	0.109	0.093	0.244

Table 4: Normal MLE Parameter Estimates

Variable	Coefficient	Standard Error
constant	1.440	0.037
male	0.116	0.008
AFQT	0.034	0.004
Mhgc	-0.003	0.001
hgc	0.040	0.003
exper	0.052	0.002
Diploma	-0.002	0.011
AA	0.134	0.015
BA	0.168	0.013
workSch	-0.014	0.001
workPT	-0.056	0.009
sigma	0.486	0.007
log likelihood	-17,167.2	

Table 5: Normal MLE Parameter Estimates with and without Gradient

(a) No User-Supplied Gradient

Variable	Coefficient	Standard Error
constant	1.427	0.037
male	0.116	0.006
AFQT	0.034	0.004
Mhgc	-0.003	0.001
hgc	0.041	0.003
exper	0.052	0.002
Diploma	-0.004	0.011
AA	0.131	0.015
BA	0.167	0.013
workSch	-0.014	0.009
workPT	-0.056	0.009
sigma	0.486	0.002
log likelihood	-17,167.32	
iterations	133	

(b) User-Supplied Gradient

Variable	Coefficient	Standard Error
constant	1.428	0.037
male	0.116	0.006
AFQT	0.034	0.004
Mhgc	-0.003	0.001
hgc	0.041	0.003
exper	0.052	0.002
Diploma	-0.003	0.011
AA	0.134	0.015
BA	0.166	0.013
workSch	-0.014	0.009
workPT	-0.056	0.009
sigma	0.486	0.002
log likelihood	-17,167.27	
iterations	49	

Table 6: Normal MLE Parameter Estimates with and without Hessian

(a) No User-Supplied Gradient or Hessian

Variable	Coefficient	Standard Error
constant	1.427	0.037
male	0.116	0.006
AFQT	0.034	0.004
Mhgc	-0.003	0.001
hgc	0.041	0.003
exper	0.052	0.002
Diploma	-0.004	0.011
AA	0.131	0.015
BA	0.167	0.013
workSch	-0.014	0.009
workPT	-0.056	0.009
sigma	0.486	0.002
log likelihood	-17,167.32	
iterations	133	

(b) User-Supplied Gradient and Hessian

Variable	Coefficient	Standard Error
constant	1.428	0.037
male	0.116	0.006
AFQT	0.034	0.004
Mhgc	-0.003	0.001
hgc	0.040	0.003
exper	0.052	0.002
Diploma	-0.003	0.011
AA	0.134	0.015
BA	0.166	0.013
workSch	-0.014	0.009
workPT	-0.055	0.009
sigma	0.486	0.002
log likelihood	-17,167.27	
iterations	45	