MA771 / MA858

Practicals and Exercises

B.J.T. Morgan

January 26, 2009

MA771 / MA858
Practical 1

The course folder is to be found at

`V:\courses\MA771.`

The printer that we use in the room for the computer practical sessions is: DPCC02.
The computer programs that we shall use exist in both MATLAB and R forms, and they are all available on the course folder. Here we shall focus on MATLAB. In order to access MATLAB through the campus network, we progress through the following sequence:

Start → All Programs → departmental software → IMSAS → MATLAB

After clicking on MATLAB you enter the program, and can access the three default screens. For explanations, see the Appendix B of the course book, starting on page 275. What you have to do is set the appropriate Directory by browsing in order to find the "m-files" directory in the courses folder. Note that at times you will need to copy programs onto your own local directory. This is so that you can edit programs before running them, and also so that you can set up files which keep records of what you do in a Practical session. You "browse" by clicking on the small rectangle at the top, with three dots :"..." in it. To start with, you should just focus on the Command window, which is the largest of the three, and type into that. The prompt in the Command window is : >>. MATLAB programs are stored in the computer as files that have names that end in: ".m". This is the way that MATLABfinds programs: it searches all the files with that suffix. So this is quite a useful device. However when we specify files we do not need to type the ".m" terms, as we shall see below.
The first practical exercise is a fairly simple one. You will construct a couple of simple likelihoods and a couple of likelihood surfaces, all of which you have already seen in lectures. You will observe that generated graphics appear in separate "Figure" windows. You will also open programs using the MATLAB editor, and simply observe the way that the programs are written.

The tasks for the first session are given below. You type the commands in the different type-face, which follow the MATLAB prompt, >>.

>> `likels` (now we do not use the ".m")

Note that you can, if you wish, edit the figure, using the tools available; for instance you could now insert axis headings, if you wish. You can also print the figure, should you want to. Now click on `likels.m` in the Current Directory window. This will cause the file to be opened in the MATLAB editor. Figure and editor windows may shrink to icons at the foot of the screen. Clicking on them opens them again. More than one Figure window may be generated by a program.

>> `betagecon`

Try opening the program file using the editor. Try changing the number of contours, and re-running the program. Note that after editing a file, you need to save it before running the program again.

>> `cauproj`

Try clicking [left mouse button] on the rotate icon,
and then clicking and dragging on the projection

>> `boha`
As for the last exercise above; this is the Bohachevsky surface

>> `quit` (NB. It is very important to quit from MATLAB
once you have finished your session.)

MA771 / MA858
Practical 2

In this session we aim to run programs to minimise functions numerically, and to check gradients.

(a) In order to appreciate better the possibilities in MATLAB click on "help". For example, look at `fminsearch` help.

There are also various demonstration programs. Try:

>> `demo` (try graphics, 3-D Plots)

(b) >>`fminbnd ('liky', 0, 1)`

Note that "fminbnd" is a MATLAB function that minimises a function over the given range. If you do not specify a large enough range, then the method will return the appropriate range end-point.

(this numerically optimises the "smokers" likelihood.

You should get answer of 0.2241 (see p20 of the course book)

>> `grad ('liky', ans)`

(This checks the value of the gradient at the maximum)

Open the file "newton.m" in the editor, and consider the code.

>> `newton` (this performs NR for the Cauchy example)

>> `global data`

Note that this makes available the data used in the last program, for use in the next.

>> `fminsearch ('cauchy', [.5, .6]);`

Note that this performs the simplex method for the two-parameter Cauchy log-likelihood)

```
>> grad ('cauchy', ans)
```

If you want to keep records you can copy files to your own filespace, and then use "diary". *Note that you will need all relevant files to be in your home directory.*

```
>> diary monday
>> fminbnd ('liky', 0, 1)
   ⋮
>> diary off
>> quit
```

MA771 / MA858
Practical 3

A. In this session we aim to use "diary" to get results, use equation (4.2) to get approximate standard errors etc., and try simulated annealing.

B. Copy, from the `m-files` folder, the following programs to your filespace:

grad.m	hessian.m	fmax.m	annealing.m
preference.m	cauchy.m	begeo.m	logit.m

C.
```
>> diary week4
>> global data  x  n  r
>> data = [1.09  - 0.23  0.79  2.31  - 0.81];
>> fmax ('cauchy', [0.05  0.63])
>> data = [29  16  17  4  3  9  4  5  1  1 1  3  7];
>> fmax ('begeo', [.5  .2])
>> fminsearch ('preference', [0  0  0  0])
>> x = [49.06  52.99  56.91  60.84  64.76  68.69  72.61  76.54];
>> n = [59  60  62  56  63  59  62  60];
>> r = [6  13  18  28  52  53  61  60];
>> fminsearch  ('logit', [.1  .2])
>> plot  (log(x), r./ n, '*')
>> hold
>>  z = 45 :  1 :  80;
>> w = ones (size (z));
>> y = w./ (1 + exp (ans (1) + ans (2) * log (z)));
>> plot (log (z), y);
>> diary off
>> quit
```

Note that we start the optimisation of the “preference” function at the $[0, 0, 0, 0]$ point because the function is parameterised on the logistic scale. This means that on the untransformed scale each parameter is started at $1/2$.

D. Try out simulated annealing. Before you can run the program, you need to make data global, and enter numbers into the data vector.

MA771 / MA858
Practical 4

1. An exercise in showing that Newton-Raphson can diverge:
   copy to your home directory the file: `newton.m`
   run it in MATLAB – it shows the progress of the iterations.
   Use the editor to change `x0`, and run again; continue until divergence.

2. Getting a profile log-likelihood:
   First you need to copy: `beopt.m`
   `begeo.m`
   `profile.m`.
   ```
   >> global data
   >> data = [29 16 17 4 3 9 4 5 1 1 1 3 7];
   >> profile
   ```

3. copy `cauproj.m`
   ```
   >> subplot (2, 2, 1)
   >> cauproj
   >> subplot (2, 2, 2)
   >> cauproj
   ```
   Now rotate it, in order to obtain a different perspective.
   ```
   >> subplot (2, 2, 3)
   ```
   ...and continue in this way.

MA771 / MA858
Practical 5

The exercises listed below relate to parameterisation and score tests.

1. Use `fmax` on two alternative parameterisations of the logit model. Copy the programs `logit.m` and `logit2.m`. The first uses $p = 1/(1+e^{\alpha+\beta x})$. The second uses $p = 1/(1+e^{\beta(x-\mu)})$.

2. Copy program `rabbits.m`.

   ```
   >> global data
   >> data = [48 20 7 5 2 2 1 2 0 0 1];
   >> fminbnd ('rabbits',                    );
                         ↑ suitable range
   ```

   List `rabbits.m` and check the code.

3. Copy `pois0.m` and `score.m`.

   ```
   >> score ([0, 0.4552])
   ```

   Why use 0.4552?

4. Copy `longscore.m` and `preference.m`.

   ```
   >> data = [29 16 17 4 3 9 4 5 1 1 1 3 9];
   >> longscore ([-10 -10 0.275 0.091])
   ```

   Note here that we use "-10" as example large negative numbers, as we are using a logistic transformation, and really want the key parameters to be zero. Note also that the numbers: 0.275 and 0.091 come from Table 4.1 on p 81 of the course book.

MA771 / MA858
Practical 6

The exercises below provide experience with the EM algorithm.

1. We run both programs in Figure 5.1 on p 141 of the course book:

   (a) >> `global data`

   >> `data =` [7 9 13 16 19 24 27 33 18 14 9 4 2];

   >> `fminbnd ('zeropo',` a, b)

   ↑ ↑

   you decide!

   (b) >> `zeropoem`

   >> `data (1)`

   Note that this provides an estimate of the missing data. How good is it here? Both programs fit the zero-truncated Poisson distribution.

2. We run both programs in Figure 5.2 on p 146 of the course book:

   (a) >> `global uncen cen`

   >> `cen =` [28 48 49];

   >> `uncen =` [1 3 3 6 7 7 10 12 14 15 18 19 22 26 29 34 40];

   >> `fminbnd ('censor',` a, b)

   ↑ ↑

   you decide!

   (b) >> `expem`

3. >> `poismix`

   Note that this program also needs the program `divide.m`.

   You may like to edit to change `nit`.

MA771 / MA858
Practical 7

The aim of this practical is to gain experience with simulation and with constructing kernel density estimates.

1. You will need to copy the following files: `gammasim.m`, `binsim.m`, `betasim.m`, `poissim.m`, `kerplot.m`, `kernel.m`, `delta.m`.

2. >> `y = rand (100, 1);`

   >> `hist (y)`

3. >> `y = gammasim (2, 200);`

   >> `hist (y)`

   >> `hold`

   >> `data = y;`

   >> `kerplot`

4. Edit `kerplot` to change k1, then re-run `kerplot`. Repeat as you wish.

5. Experiment with `betasim`, etc. Try to relate results to known features of distributions, such as means. Check program listings for format.

MA771 / MA858
Practical 8

The aim of this practical is to obtain bootstrap and parametric bootstrap samples, to summarise these using a kernel density estimate, and to run Gibbs sampling and Metropolis Hastings sampling for a simple example.

1. You will need to copy the following files: `paraboot.m, boot.m begeo.m, multsim.m, gibbs.m, betasim.m, binsim.m, metro.m, biv.m`

2. >> `paraboot`

   >> `h=findobj(gca, 'Type', 'patch');`

   >> `set (h, 'FaceColor', 'y')`

   Display figure 2, by typing: `figure(2)`. Note that here we focus on theta

   >> `data = x2;`

   >> `hold`

   >> `kerplot`

   >> `mean (x2)`

   >> `var (x2)`

3. Repeat above for "`boot`"

4. >> `gibbs;`

   >> `t = 1:k;`

   >> `plot (t, x1)`

5. >> `metro`

MA771 / MA858
Exercise 1

The numbers below relate to the Exercises in the course book.

1. 2.2 Consider both the geometric and the beta-geometric models.

2. 2.3.

3. 2.15 (i) (a).

MA771 / MA858
Exercise 2

1. 3.10: use the data of Practical 3, C. Note that in order to plot the data, use: >> `plot (log(x), r./ n, '*')`.

2. Use the annealing program in file: `annealing.m` and experiment with simulated annealing for maximising the Cauchy log-likelihood:

   >> `annealing ('cauchy', [0.5, 0.6]).`

   Note that you first have to make data global, and then enter the data:

   >> `data=[1.09.  etc.]`

3. 3.12.

4. 3.13.

5. 3.15, but for the case where each censored individual has his/her own censoring time.

MA771 / MA858
Exercise 3

1. Provide an example of the divergence of the Newton-Raphson method, and comment on your results.

2. Provide an example of a profile log-likelihood, and discuss your results.

3. Repeat the model selection of Table 4.3 on page 97 of the course text, but using the BIC, rather than the AIC, for the "Other contraceptive users".

4. The solution to Exercise 4.8 on p 296 of the course text provides some of the detail of the score

   test algebra of Example 4.10 on p 103 of the course text. Form $\frac{\partial^2 \ell}{\partial\alpha\partial\lambda}$, and if possible, $\mathbb{E}\left[\frac{\partial^2 \ell}{\partial\alpha\partial\lambda}\right]$.

MA771 / MA858
Exercise 4

1. Write up results of Practical 5, Q1, and discuss.

2. Write up results of Practical 5, Q2. Obtain the maximum-likelihood estimate of the two parameters of the zero-inflated Poisson distribution for the rabbit data $\widehat{w}$ and $\widehat{\lambda}$.

3. Perform a score test of the hypothesis that $w = 0$ in the zero-inflated Poisson distribution. How does the result compare with a likelihood-ratio test? (In order to do a LR test, we need to maximise the likelihood with $w = 0$, and when $w \neq 0$. In the latter case, in MATLAB , we need: $>>$ `pois0` ([mle]).

4. Write up the results of Practical 6.

5. Explain how the program of Figure 5.2(b) works.

MA771 / MA858
Exercise 5

1. Use the inversion method to simulate from the following distributions:

   (a) Pareto: $\Pr(X \leq x) = 1 - \left(\frac{k}{x}\right)^a$
   for $a > 0,\ x \geq k > 0.$

   (b) Extreme-value distribution:

   $$\Pr(X \leq x) = \exp\{-\exp((\xi - x)/\theta)\}, \quad \text{for } \ x \geq 0.$$

2. Check that the naive density estimate is different from the traditional histogram.

3. 6.13. The data here are in the data directory, in ex6.13.dat

4. 6.16. Here use the `ecosim` program.