Advanced Programming in Quantitative Economics

Introduction, structure, and advanced programming techniques

Charles S. Bos

VU University Amsterdam
Tinbergen Institute

c.s.bos@vu.nl

15 – 19 August 2011, Aarhus, Denmark

Tutorial Day 2 - Afternoon

14.30P Tutorial

- Exercises HB
- Data duration model

16.00 End

Rows or columns

Does it make a difference to use rows or columns, for the speed of your program? Find this out...

- Make a program which fills a `n x n` matrix with random numbers, e.g. with $n = 1000$
- Do this either
  1. at once, using `rann(iN, iN);`
  2. filling mX row by row, `mX[i][]= rann(1, iN);`
  3. filling mX column by column, `mX[][i]= rann(iN, 1);`
- Measure the time each of these takes. Use the `dTime= timer();` and `println (timespan(dTime));` to measure.
- If your computer is too quick, repeat the filling say $S = 10$ or $S = 1000$ times, or more, until you see which one works better. Or increase n?

Precision/smallest number in Ox

What is the smallest number you can create which is still larger than zero?

1. Start of with a number of choice
2. Check if it is different from 0
3. Divide it by two
4. and repeat from 2

Report the last number which you found different from zero. Also report the number of times you divided by 2.
For this exercise you might want to use a construction like

```
do
  {
    // Do something}
  }
while (dX != 0);          // Test whether dX is equal to zero
```

Accuracy in addition

Define three numbers

$$a = 0.1234567 \times 10^0 \qquad b = 0.4711325 \times 10^4 \qquad c = -b$$

and compute the outcomes of

$$a + b + c$$
$$a + (b + c)$$
$$(a + b) + c$$

Is there a difference?
Afterwards, do the same thing but with 10^{40} instead of 10^4: Do you now find a difference? Can you find the number of significant digits With what k, for 10^k, does the result seem correct?

Memory use

Does declaring new memory, or using local variables, take time? Investigate this by

1. Writing a loop which S times creates a random matrix of size $n \times n$ of random numbers, with n, S sufficiently large, and time it
2. Then do the same, but each time in the loop reset the matrix to a scalar 0, before assigning the big matrix to it.
3. Then do the same, but call a function which locally declares a matrix `mX` and assigns the random numbers to it.

Any difference? (Hint: I cannot find much of a difference myself)

Duration modelling

The duration model is heavily used e.g. to model the duration of unemployment spells. It also provides a convenient workhorse during this course as it

- contains relatively few parameters
- does have restrictions on the parameter space
- can be estimated using a loglikelihood approach
- allow for easy extension from the regression framework

See a.o. Lancaster, 'The Econometric Analysis of Transition Data' (1990) for details.

Here we use a simplified version of the model, assuming all data is observed.

Duration: The (simplified) model

Durations y_i are assumed to be distributed according to

$$y \sim \text{Weib}(\alpha, \lambda) \qquad f(y; \alpha, \lambda) = \alpha \lambda^{\alpha} y^{\alpha - 1} \exp(-\lambda^{\alpha} y^{\alpha})$$

Dependence on personal characteristics can be introduced by taking

$$\lambda_i \equiv \exp(X_i \beta)$$
$$y_i \sim \text{Weib}(\alpha, \lambda_i)$$

Duration: Simulation

Write a program which generates $N = 1000$ durations y from the Weibull model, with $\beta = (1\ 1)', X = [1\ N(0,1)], \alpha = 1.5$.
Some remarks:

- Think about the status of $y, X, \lambda, \alpha, \beta$: Which is parameter, which is 'fixed data', 'derived data' etc?
- Work *in matrices* as far as possible.
- Work in routines: In what steps can you generate this data? What should you retain?
- If you generate $E \sim \text{Exp}(\lambda_e = 1)$, then $Y = E^{1/\alpha}/\lambda_w \sim \text{Weib}(\alpha, \lambda_w)$. Use this relation, and the element-by-element division and power operators `./` and `.^` to obtain a sample from the requested Weibull.
- `ranexp` does not work immediately. You'll miss a line `#include <oxprob.h>` at the beginning. What is the logic?

Duration: Output

In later tutorials, we will need data from this model. Therefore,

- Save a data set `data/genrdur.fmt` containing for each individual the duration y and the explanatory variables X.
- Get some output on the y's and X's.

Don't try to do all at once: First check that you can generate e.g. from the Weibull for fixed λ, then generate separate λ's, etc...