MAS115 R programming 2016-17

Using pseudocode to write computer code: Wind farms

In this handout I want to show how easy it is to write computer code **ONCE** you have got good pseudocode. Remember our question about locating a windfarm and how many days we'd have low wind speeds.

1 Problem:

A particular site is being considered for a wind farm. At that site, Y_t, the log of the wind speed in m/s on day t is known to depend upon the previous two days winds:

$$Y_t = 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t, \tag{1}$$

with $\varepsilon_t \sim N(0, 0.01)$. If $Y_1 = Y_2 = 1.5$, what is the probability that the wind speed $\exp(Y_t)$ will be below 4 m/s for more than 10 days in a 100 day period?

Solution style - Monte Carlo estimation

- Create n potential series of 100 day wind speeds,
- Find out the proportion with less than 10 days of low wind,
- If n is large this will be a good estimate of probability.

Coding

In this handout, we show the pseudocode and then give R code which will implement it. I have given two versions of code:

1. A version using two nested `for` loops — this is the easiest method to understand initially as it is a verbatim translation of the pseudocode. Look at this first and make sure you understand it. It is however a bit slow to run, as explicit loops often slow down your computer when running the code.
2. A version using vectorisation i.e. the fact that R, like some other languages, allows you to create/operate on whole vectors and matrices jointly rather than having to identify each element separately. This is much faster to run.

The second version is better from a programming point of view as it is much quicker (something which will often be important). If you can understand it (and use it in your own programming) then great but **don't worry** if, for the moment, you program using the first approach. It's the ideas which are more important and, as you practice more, you'll get more used to vectorising things.

2 Pseudocode for the solution

2.1 Creating a single realisation of the next 100 days

We're breaking this up one part at a time. First let's create a single hypothetical set of wind speeds for the 100 days.

1. CREATE Y as vector of length 100
2. Set Y[1] = Y[2] = 1.5
3. FOR ($t = 3, 4, \ldots, 100$):
   - Sample ε_t from $N(0, 0.01)$
   - Set $Y_t \leftarrow 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$

   ENDFOR
4. Count number of elements of $\{Y_1, \ldots, Y_{100}\}$ less than $\log 4$:
   - Set $X_i \leftarrow \sum_{t=1}^{100} \mathbb{1}_{[Y_t < \log 4]}$
5. Determine if event E has occurred for time series i:
   - Set $E \leftarrow \mathbb{1}_{[X_i > 10]}$

2.2 Full solution: Embedding this to create lots of hypothetical futures

Now we can create a single realisation and see if it had low-wind speeds. We can embed this in another `FOR` loop to create N hypothetical sets of 100 day wind speeds and use Monte Carlo to estimate the probability. We will define E to be the event that in 100 days the wind speed is below 4 m/s for more than 10 days. To estimate $P(E)$, generate lots of individual time series, and count proportion of series in which E occurs.

- FOR ($i = 1, 2, \ldots, N$):
  1. Generate ith realisation of the time series process:
     CREATE Y as vector of length 100
     Set Y[1] = Y[2] = 1.5
     FOR ($t = 3, 4, \ldots, 100$):
     - Sample ε_t from $N(0, 0.01)$
     - Set $Y_t \leftarrow 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$

     ENDFOR
  2. Count number of elements of $\{Y_1, \ldots, Y_{100}\}$ less than $\log 4$:
     - Set $X_i \leftarrow \sum_{t=1}^{100} \mathbb{1}_{[Y_t < \log 4]}$
  3. Determine if event E has occured for time series i:
     - Set $E_i \leftarrow \mathbb{1}_{[X_i > 10]}$

  ENDFOR
- Estimate $P(E)$ by $\frac{1}{N} \sum_{i=1}^{N} E_i$

3 Using the pseudo code

3.1 Line by line implementation of a single future

Pseudocode

1. CREATE Y as vector of length 100
2. Set Y[1] = Y[2] = 1.5
3. FOR ($t = 3, 4, \ldots, 100$):
   - Sample ε_t from $N(0, 0.01)$
   - Set $Y_t \leftarrow 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$

   ENDFOR
4. Count number of elements of $\{Y_1, \ldots, Y_{100}\}$ less than $\log 4$:
   - Set $X_i \leftarrow \sum_{t=1}^{100} \mathbb{1}_{[Y_t < \log 4]}$
5. Determine if event E has occured for time series i:
   - Set $E \leftarrow \mathbb{1}_{[X_i > 10]}$

R code

```
y[1] <- y[2] <- 1.5              # Initialise first two values
for(t in 3:100){                 # For loop to sample next day
  y[t] <- 0.6*y[t-1]+0.4*y[t-2 ]+rnorm(1,mean = 0,sd = 0.1)
}
nlow <- sum(y<log(4))            # How many wind speeds < 4 (log as Y=log wind)
E <- (nlow > 10)                 # Are there less than 10?
```

3.2 Using the pseudo-code - lots of time series

Pseudocode

- FOR ($i = 1, 2, \ldots, N$):
  1. Generate ith realisation of the time series process:
     CREATE Y as vector of length 100
     Set Y[1] = Y[2] = 1.5
     FOR ($t = 3, 4, \ldots, 100$):
     - Sample ε_t from $N(0, 0.01)$
     - Set $Y_t \leftarrow 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$

     ENDFOR
  2. Count number of elements of $\{Y_1, \ldots, Y_{100}\}$ less than $\log 4$:
     - Set $X_i \leftarrow \sum_{t=1}^{100} \mathbb{1}_{[Y_t < \log 4]}$
  3. Determine if event E has occured for time series i:
     - Set $E_i \leftarrow \mathbb{1}_{[X_i > 10]}$

  ENDFOR
- Estimate $P(E)$ by $\frac{1}{N} \sum_{i=1}^{N} E_i$

R code

```
N <- 1000
E <- rep(NA, length = N)
for(i in 1:N) {
  y <- rep(1.5,100)        # May as well just initialise vector with value 1.5
  for(t in 3:100){
     y[t] <- 0.6*y[t-1]+0.4*y[t-2]+rnorm(1, 0, 0.1)
  }
  nlow <- sum(y<log(4))
  E[i] <- (nlow > 10)
}
mean(E)
```

4 **Pretty R code using vectorisation - Wind Farms**

```
# Pretty solution to wind farm using vectorisation
# Generate a multiple time series, count number of series
# in which event E occurs

n <- 1000
y<-matrix(1.5,100,n)
noise<-matrix(rnorm(98*n,0,0.1),98,n)
for(i in 3:100){
  y[i,]<-0.6*y[i-1,]+0.4*y[i-2,]+noise[i-2,]
}
nlow<-apply((y<log(4)),2,sum)
mean(nlow>10)
```