MAS115: R programming
Lecture 3: Monte-Carlo Estimation
Lab Class: more loops via `while` and `repeat`

Bryony Moody

February 23, 2021

Aims

In lab we will learn two more methods of creating loops in R

- `while`
- `repeat`

Monte-Carlo methods:

- Formalisation
- Wind Farms

Demonstrate how to use our pseudocode to write our program step-by-step

Monte Carlo simulation

Last time we looked at approximating the *expectation* $\mathbb{E}[X]$ of a random variable X by a sample mean

$$\frac{1}{n}\sum_{i=1}^{n} X_i.$$

More generally, if we want to estimate $\mathbb{E}[h(X)]$ for any function $h(\cdot)$, we can use

$$\mu = \mathbb{E}[h(X)] \approx \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

if the $\{X_i\}$ form a *large* sample from the density $f_X(\cdot)$ of X.

Monte Carlo simulation

In fact—extending what we saw last time—it can be shown that, for large n,

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Our Monte Carlo estimate will be centred on the true value, with a variance around it that decreases with n in a well-understood way.

Monte Carlo simulation

An important special case is when $h(\cdot)$ is an *indicator function*, taking the value 1 when some event involving X occurs, and 0 otherwise.

Some times written $\mathbb{1}_E$ where E is the event.

For example, we can estimate the probability that $X < 0$ by taking $h(X) = \mathbb{1}_{\{X<0\}}$;

$$p = \Pr\{X < 0\} \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i<0\}}.$$

Monte Carlo integration

The definition of an expectation is actually an integral.

Often, the value I of a definite integral can be written as the mean of a function of a random variable—even though I is not random. Then we can approximate the integral:

- get our computer to sample lots of values of that random variable;
- work out the sample mean.

Monte Carlo integration

Say

$$I = \int_a^b f(x)\,\mathrm{d}x.$$

If $X_1, \ldots, X_n \sim U[a, b]$, then

$$\mathbb{E}f(X_i) = I/(b-a),$$

and so

$$I \approx \frac{(b-a)}{n} \sum_{i=1}^{n} f(X_i).$$

Monte-Carlo Simulation: Pseudocode to Computer Code

Situating a wind farm

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:

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

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?

Situating a wind farm

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:

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

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 series of wind speeds.
- Find out the proportion with more than 10 days of low wind.
- If n is large this will be a good estimate of probability.

Situating a wind farm - Single Realisation

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

Initial Task:
Write pseudocode which generates a vector Y which

- Records a hypothetical set of log-wind speeds for 100 days with starting values $Y_1 = Y_2 = 1.5$
- Finds if there have been more than 10 days of winds below 4 m/s. Store this as

$$E = \begin{cases} 1 & \text{if more than 10 days below 4 m/s} \\ 0 & \text{if 10 or fewer days below 4 m/s} \end{cases}$$

Single Realisation Wind-Farm Pseudocode

CREATE Y as vector of length 100
Set $Y_1 = Y_2 = 1.5$

1. 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 occurred for time series i:
   - Set $E \leftarrow \mathbb{1}_{[X_i > 10]}$

Situating a wind farm - Full solution

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.

Define E: 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.

Pseudocode to solve the whole problem:

- 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 occurred for time series i:
     Set $E_i \leftarrow \mathbb{1}_{[X_i > 10]}$

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