
Practical 1: Turning Algorithms into code



University of
St Andrews

Contents

Some background: wind power	2
Getting started: the data	4
Rejection sampler	6
Monte Carlo error	7
Inverse CDF sampler	8
Comparing the samplers	9
How many samples?	10

Some background: wind power

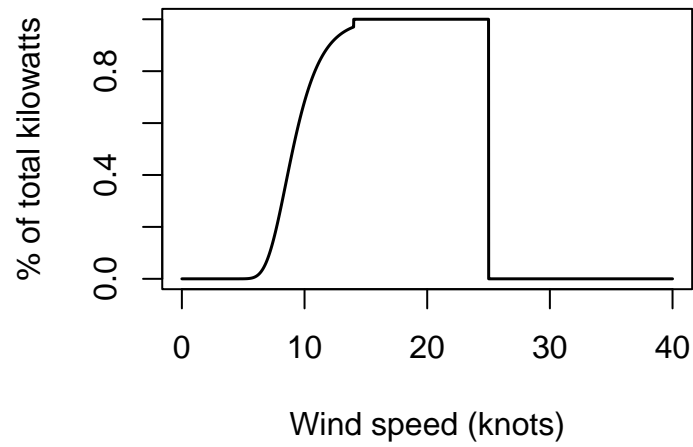
A company is considering whether to install a wind turbine at a particular site in the UK. The company wishes to estimate how much wind power would be generated at the site. There are also characteristics of the site that are of interest so that the company can decide what type of turbine to install. To answer these questions, the company needs you to develop software that can simulate wind speed and the power produced by these wind speeds.

For a given wind speed, w in knots, the proportion of kilowatt power produced, $P(w)$ (as a proportion of the maximum power the turbine can produce), is given by the mathematical function:

$$P(w) = \begin{cases} 0 & \text{if } w < 3.5; \\ \exp(-213.974 \exp(-0.633w)) & \text{if } 3.5 \leq w \leq 14; \\ 1 & \text{if } 14 < w \leq 25; \\ 0 & \text{if } w > 25. \end{cases}$$

The power function is zero until there is enough wind to turn the turbine (at 3.5 knots). The power then increases until it reaches the maximum that the turbine can produce. At 25 knots, it is too dangerous to use the turbine, so it produces no power.

Wind turbine power output



Exercise 1

Write a function that computes the power for a given vector of wind speeds. Use this function to create a plot similar to the one displayed above.

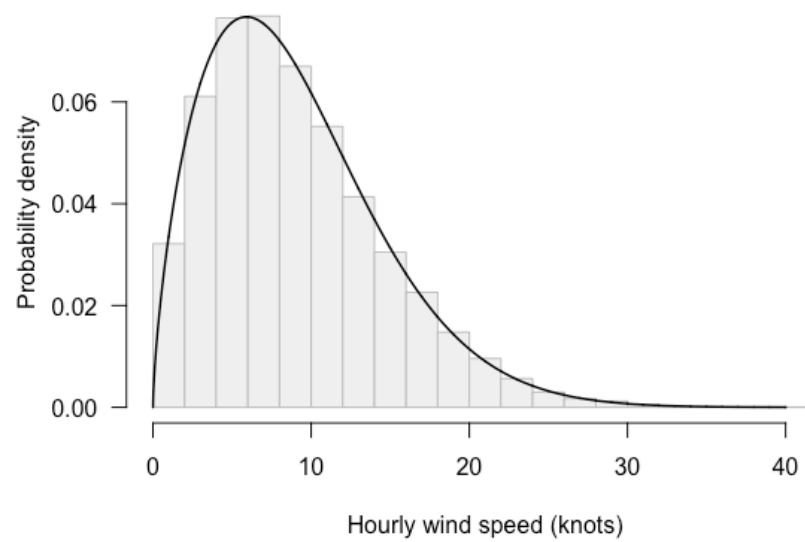
Getting started: the data

At the proposed site, the wind speed was recorded hourly for approximately 440 days. The data are stored in a comma separated value file (.csv), which is available on Moodle. Download it into your working directory. Load it into R using the following command:

```
dat <- read.csv("WindData.csv")$speed
```

The climatologists involved decide that wind speed is best modelled using a Weibull distribution. The Weibull distribution has two parameters: a shape parameter k and a scale parameter λ . The probability density function can be computed using the `dweibull` function in R. The climatologists compute

the following estimates of these parameters for these data: $\hat{k} = 1.679$ and $\hat{\lambda} = 10.128$.



Exercise 2

Use the `dweibull` function and the parameter estimates to produce a plot similar to the one displayed above.

Rejection sampler

In Chapter 2, the algorithm for the rejection sampler was given, which works as follows. Repeat until you have n samples:

1. Pick a random coordinate x .
2. Pick a random coordinate y .
3. Compute the probability density at x , $f(x)$.
4. If $y \leq f(x)$, then accept x .

Exercise 3

(3a) Write a function that has inputs n, k , and λ and outputs n samples from the Weibull distribution using a rejection sampling method.

(3b) Using your sampler, write code to answer the following question: what is the mean hourly power that could be produced at the site?

Monte Carlo error

When you report the mean hourly power produced, the company ask how accurate that estimate is. Monte Carlo methods are simulation-based methods, so every time you estimate the mean power you will get a different answer. The company wants to know how much that answer changes. One way to answer that is to compute the mean many times and quantify how much it changes. Repeat m times:

1. Generate n samples using rejection sampler function $f(n, k, \lambda)$.
2. Compute power produced by each wind speed.
3. Compute mean power produced.
4. Compute standard deviation of mean power.

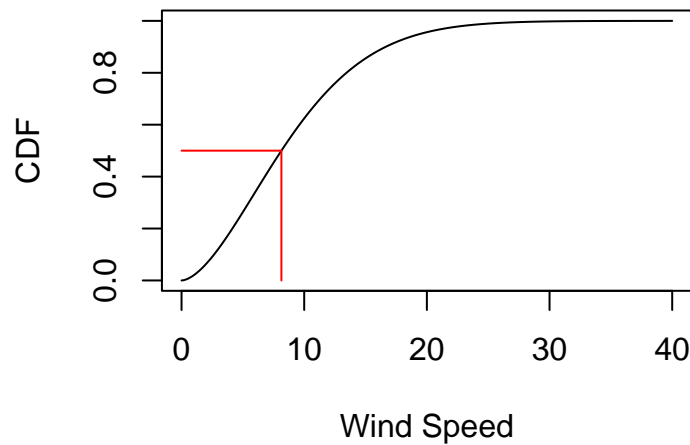
In practice, it is usually not feasible to simulate multiple times in order to quantify Monte Carlo error, so the following approximation is used: σ/\sqrt{n} , where σ is the standard deviation of the samples and n is the number of samples.

Exercise 4

Write a function with inputs m, n, f, k , and λ , and which outputs the standard deviation of the mean power from n samples.

Inverse CDF sampler

An alternative method is proposed for sampling from the Weibull distribution; it is called the inverse cumulative distribution function (CDF) sampler or sometimes called the probability integral transform. The idea of this method is to use the cumulative distribution function:



In this method, you pick a random number of the y -axis (between 0 and 1), then use the inverse cumulative distribution function `qweibull` to figure out the corresponding value on the x -axis. You take this value as a sample. Notice, there is no rejection here, every iteration leads to a sample.

Exercise 5

(5a) Write out in pseudo-code the algorithm for the inverse CDF sampler. What are the inputs, what are the outputs? \ (5b) Write a function that implements this algorithm. \ (5c) Use this sampler to estimate the mean power produced at the site.

Comparing the samplers

The methods can be compared by their Monte Carlo error and by how long it takes to generate samples. The time it takes to execute a statement can be computed using the function `system.time`, e.g.:

```
system.time(rnorm(10000))  
  
##      user  system elapsed  
## 0.001    0.000    0.001
```

Exercise 6

Compute the Monte Carlo error for each sampler and the time it takes to generate 10000 samples for each. Write a sentence comparing the methods and suggest which one should be used.

How many samples?

The company decides to use the inverse CDF sampler as that is the one they are familiar with. They want to know how many samples they need to take for a Monte Carlo error of 0.05. A vector of possible sample sizes can be created using the command:

```
nsamp <- seq(5, 100, 5)
```

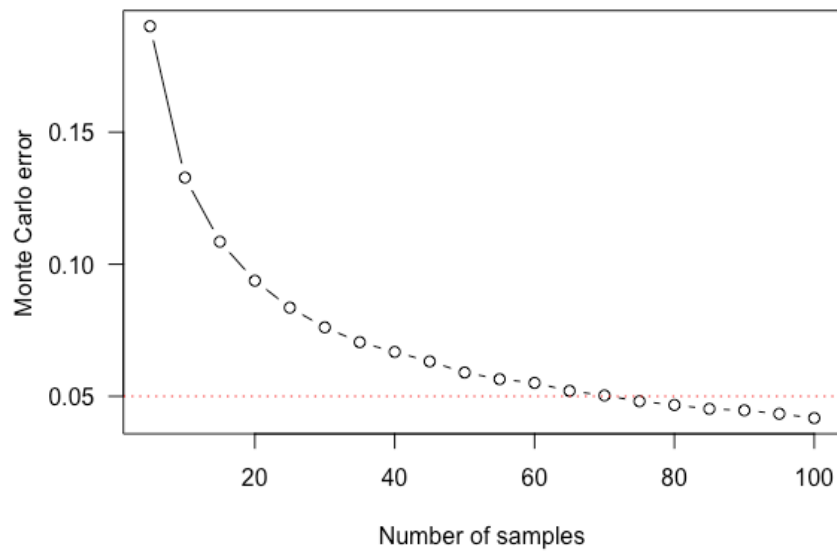


FIGURE 1 Monte Carlo Error

Exercise 7

Write an algorithm to compute the Monte Carlo error for each sample size and store it. Implement this algorithm and create a plot similar to the one above.

Finally

The company has three questions about the proposed wind turbine site:

1. What is the total daily average power that could be produced at the site?
2. What percentage of the time is the wind speed too low to produce power?
3. What percentage of the time does the wind speed produce maximum power?

To help you answer these questions, below is the code to answer the question: how often does the wind speed exceed the safe value of 25 knots? Suppose `samples` is a vector of wind speeds:

```
unsafe <- samples > 25  
percentage_unsafe <- 100 * mean(unsafe)
```

Exercise 8

Using the inverse CDF sampler, give an answer to each of the company's questions along with the code to produce these answers. For the first question, the company may ask for a range of plausible values for the average daily power produced, how could you provide this? (Hint: use the `quantile` function.)