Lab1

Anastasia Piadi, Nazli Bilgic

2023-11-06

Question 2:

a) Write your own R function, myvar, to estimate the variance in this way.

```
## myvar function with the example data(c(1,2,3,4)): 1.666667
```

```
## var function with the example data(c(1,2,3,4)): 1.666667
```

b) Generate a vector x = (x1, . . . , x10000) with 10000 random numbers with mean 108 and variance 1.

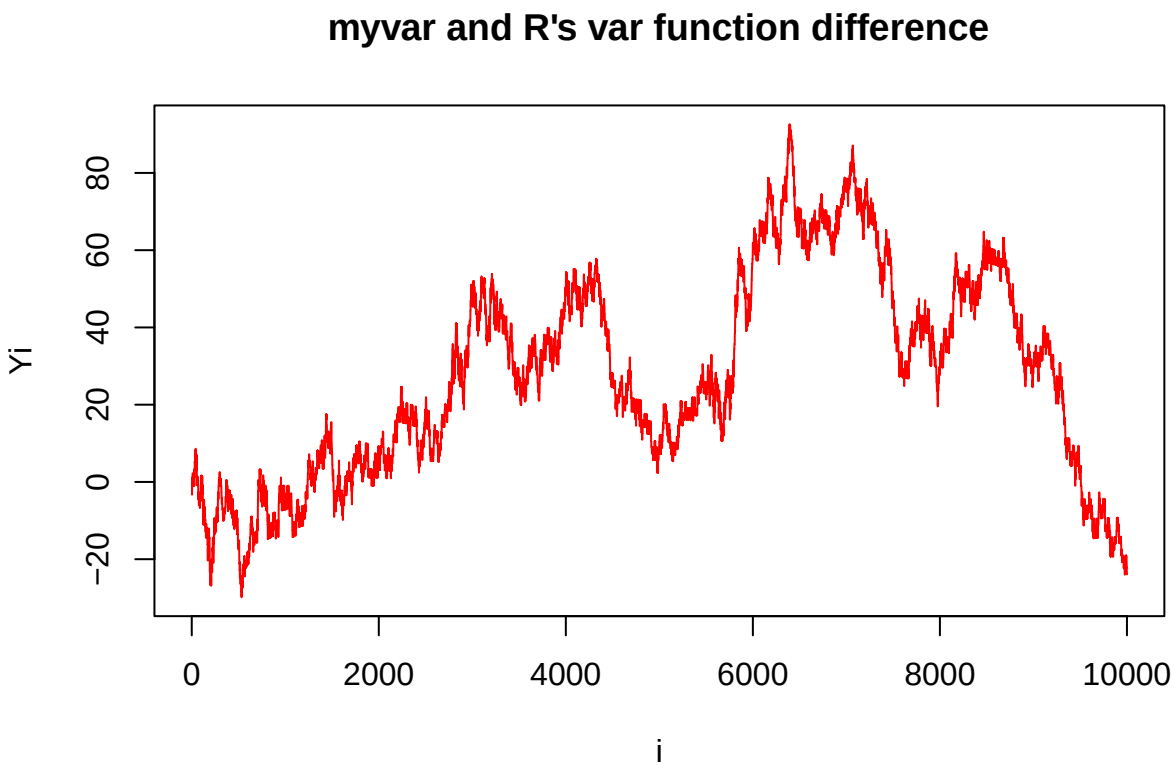
```
## myvar function output with random data: -22.93989
```

```
## var function output with random data: 0.9972751
```

c) For each subset Xi = {x1,. . . ,xi}, i = 1,. . . ,10000 compute the difference Yi = myvar(Xi) - var(Xi), where var(Xi) is the standard variance estimation function in R. Plot the dependence Yi on i. Draw conclusions from this plot. How well does your function work? Can you explain the behaviour?

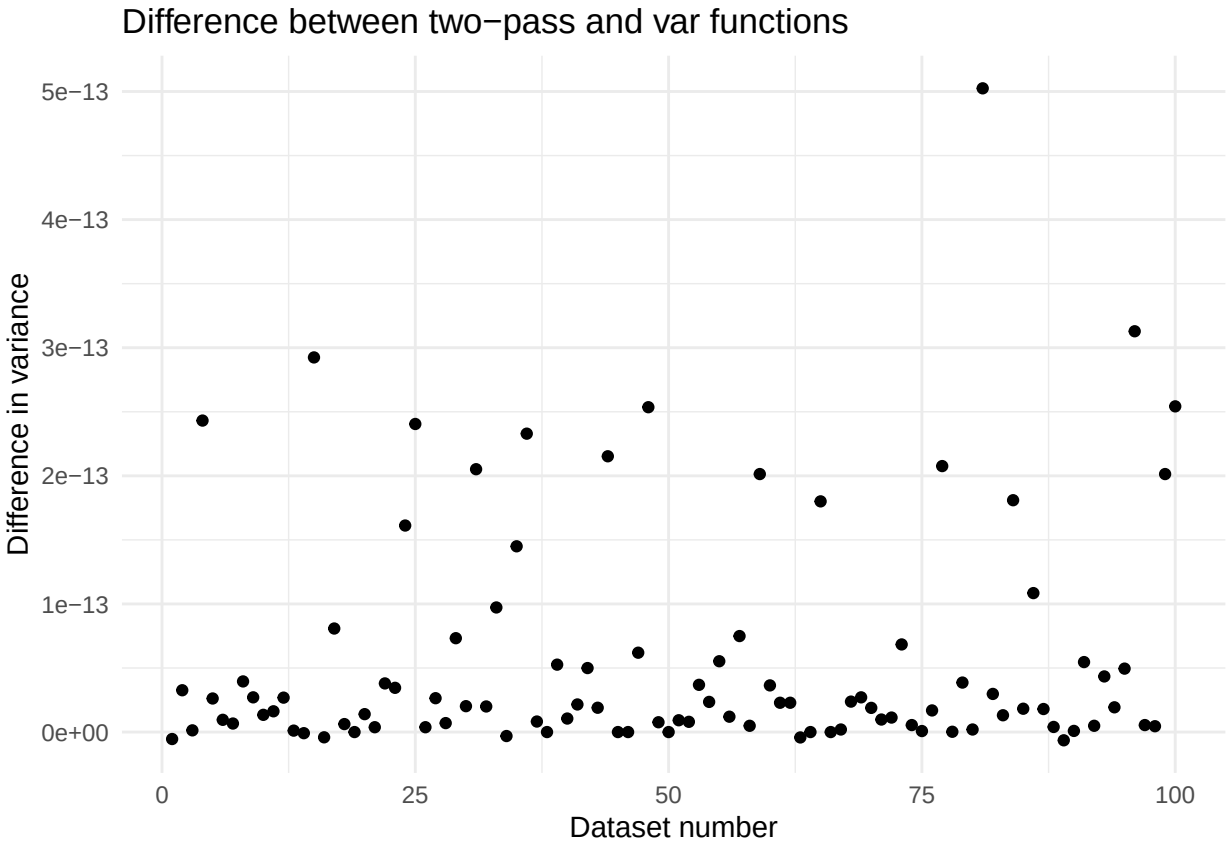
The difference between myvar and var function can be attributed to the way floating point arithmetic works in computers. This can lead to precision errors, especially when dealing with a combination of very large and very small numbers.

source: https://en.wikipedia.org/wiki/Floating-point_arithmetic#IEEE_754:_floating_point_in_modern_computers



d) How can you better implement a variance estimator? Find and implement a formula that will give the same results as var().

We used two-pass method to calculate variance in another way. In two-pass method, we first compute the sample mean then the sum of the squares of the differences from the mean. At the end, we compare the results that we get with the random vector from both methods. The results from the two-pass and var function are same. To conclude, we get better result with two-pass method fro this example.


Difference between two−pass and var functions
Difference in variance
5e−13
4e−13
3e−13
2e−13
1e−13
0e+00
0
25
50
75
100
Dataset number