

# Lab1

Anastasia Piadi, Nazli Bilgic

2023-11-05

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  $x = (x_1, \dots, x_{10000})$  with 10000 random numbers with mean 108 and variance 1.

The formula for variance in myvar function has subtraction process. When we have two large and nearly equal numbers this causes differences with the calculation. For this example we got a negative number for the variance calculation.

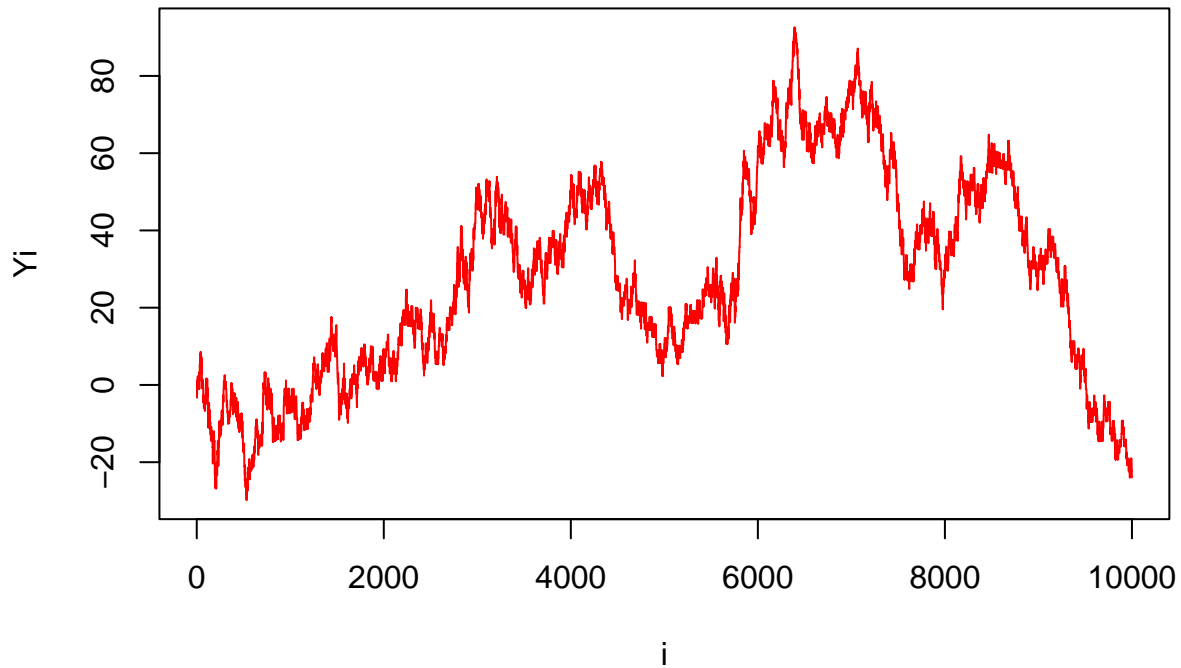
```
## myvar function output with random data: -22.93989
```

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

- c) For each subset  $X_i = \{x_1, \dots, x_i\}$ ,  $i = 1, \dots, 10000$  compute the difference  $Y_i = \text{myvar}(X_i) - \text{var}(X_i)$ , where  $\text{var}(X_i)$  is the standard variance estimation function in R. Plot the dependence  $Y_i$  on  $i$ . Draw conclusions from this plot. How well does your function work? Can you explain the behaviour?

From the graph ( $Y_i$ =difference between myvar and var func.,  $i$ =size of subset) we can see that, as  $i$  increases the difference between myvar and var function changes. This shows that the estimations vary. Difference start to get positive after  $i=2000$ , myvar function start to produce higher variance estimates then the var function. around  $i=8000$  the difference start to decrease, the myvar function starts to give close results to the var function.

### myvar and R's var function difference



- 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.

```
## Var function output: 0.9972751
```

```
## Two-pass variance output: 0.9972751
```