Project 2
Computational Statistics

The code for this project is available under
https://github.com/max607/computational-statistics-em.

By
Maximilian Schneider

26.01.2023

Contents

1 Maximum Likelihood estimation of θ

The pdf is

$$f(y_i;\theta) = \frac{\theta^2}{\theta+1}(1+y_i)\exp(-\theta y_i), i = 1\ldots n. \tag{1.1}$$

The log likelihood is

$$\ell(\theta) \propto 2n\log(\theta) - n\log(\theta+1) - \theta\sum_{i=1}^{n} y_i. \tag{1.2}$$

Setting the derivative to zero leads to equation (1.3) which has to be solved for θ.

$$\frac{\theta+2}{\theta(\theta+1)} = \bar{y}, \tag{1.3}$$

where $\bar{y}$ is the sample mean of $\boldsymbol{y}$.

2 Estimation of standard error

The pdf can be restated as

$$f(y_i;\theta) = \frac{\theta}{\theta+1}\theta\exp(-\theta y_i) + \frac{1}{\theta+1}\theta^2 y_i\exp(-\theta y_i), \tag{2.1}$$

i.e., a mixture of two gamma distributions in shape and rate parameterization. θ is the rate and the shapes are equal to 1 and 2.

It is straight forward to simulate from this, but starting from $U \overset{iid}{\sim} U(0,1)$ exponentially distributed variables can be obtained via inversion

$$f^{-1}(u;\theta) = -\frac{\log(u)}{\theta}, \tag{2.2}$$

which already is a gamma with shape one and a gamma with shape 2 is obtained via the sum of 2 gammas with shape one. For optimizing computation time n observations are generated in the following way:

1) Draw the number of shape 2 gammas (n2) by counting the number of $u < \frac{1}{\theta+1}$
2) Sample n and $n2$ uniforms
3) Transform the uniforms to exponentials using $f^(- 1)$
4) Add to the fist $n2$ of the n exponentials the other exponentials
5) Return

The result is a sample with $n2$ observations of a gamma distribution with shape 2 and $n - n2$ observation of a gamma distribution with shape 1.

For the purpose of estimating θ with the estimator of section 1 it is of no importance that the sample is sorted by shape.