Computational Statistics
Project 2

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

By
Maximilian Schneider

31.01.2023

Contents

1 Maximum Likelihood estimation of θ

The pdf is

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

The log likelihood is

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

$$\propto 2\log(\theta) - \log(\theta+1) - \theta\bar{y} + c. \tag{1.3}$$

The first derivative is

$$\ell'(\theta) = \frac{2}{\theta} - \frac{1}{\theta+1} - \bar{y}, \tag{1.4}$$

where $\bar{y}$ is the sample mean of $\boldsymbol{y}$. Note, we can drop the n.
Setting the derivative to zero leads to equation (1.5) which has to be solved for θ.

$$\frac{\theta+2}{\theta(\theta+1)} = \bar{y}. \tag{1.5}$$

One approach is to use Newton-Raphson, which requires the second order derivative.

$$\ell''(\theta) = -\frac{2}{\theta^2} + \frac{1}{(\theta+1)^2}, \tag{1.6}$$

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 is the same as a gamma with shape one and a gamma with shape 2 is obtained via the sum of 2 exponentials. 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.

3 EM

Consider the more complex pdf

$$f(y_i;\theta,\lambda,\pi) = \pi\frac{\theta^2}{\theta+1}(1+y_i)\exp(-\theta y_i) + (1-\pi)\lambda\exp(\lambda - y_i), \tag{3.1}$$

$$y_i,\theta,\lambda \in \mathbb{R}^+, \pi \in [0,1]. \tag{3.2}$$

The goal is to estimate θ, λ and π applying EM. Because (3.1) is a mixture of two PDFs, one can rewrite the observed PDF introducing missing data

$$\mathcal{L}_i(\theta,\lambda|x_i,y_i) = f_\theta(y_i)^{x_i} f_\lambda(y_i)^{1-x_i}, \tag{3.3}$$

where $X_i \sim Ber(\pi)$. We continue by applying the Bayes theorem

$$f(x_i|\theta,\lambda,y_i) = \frac{\mathcal{L}_i(\theta,\lambda|x_i,y_i)f(x_i)}{f(\theta,\lambda)}. \tag{3.4}$$

Note that $f(\theta,\lambda) = f(x_i = 0)\mathcal{L}_i(\theta,\lambda|x_i = 0, y_i) + f(x_i = 1)\mathcal{L}_i(\theta,\lambda|x_i = 1, y_i)$. Via substituting we arrive at

$$f(x_i|\theta,\lambda,y_i) = \frac{f_\theta(y_i)^{x_i} f_\lambda(y_i)^{1-x_i}\pi^{x_i}(1-\pi)^{1-x_i}}{(1-\pi)f_\lambda(y_i) + \pi f_\theta(y_i)}. \tag{3.5}$$

From this we can read

$$\mathbb{E}(X_i|\theta,\lambda,y_i) = \frac{f_\theta(y_i)\pi}{(1-\pi)f_\lambda(y_i) + \pi f_\theta(y_i)}. \tag{3.6}$$

Now we have everything we need: a approachable likelihood for the maximization step and the expectation of the missing data given the parameter estimates. All parameters are independent, so the relevant parts of the likelihoods are

$$\ell(\pi|\boldsymbol{x}) = \log(\pi)\sum_{i=1}^{n} x_i + \log(1-\pi)\sum_{i=1}^{n}(1-x_i) \tag{3.7}$$

$$\ell(\lambda|\boldsymbol{x},\boldsymbol{y}) = \log(\lambda)\sum_{i=1}^{n}(1-x_i) - \lambda\sum_{i=1}^{n}(1-x_i)y_i + c \tag{3.8}$$

$$\ell(\theta|\boldsymbol{x},\boldsymbol{y}) = 2\log(\theta)\sum_{i=1}^{n} x_i - \log(\theta+1)\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i y_i + c \tag{3.9}$$

$$\propto 2\log(\theta) - \log(\theta+1) - \theta\tilde{y} + c \tag{3.10}$$

the augmented data contains all the information about π. So, noting the distribution of X_i, we can maximize (3.7) to obtain the ML estimate $\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

It is also possible to solve (3.8) analytically. The likelihood is maximized by $\hat{\lambda} = (n - \sum_{i=1}^{n} x_i)/(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i y_i)$. $\tilde{y} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i$

4 Old

As all observations are independent we formulate the complete likelihood as

$$\mathcal{L}(\theta, \lambda, \pi|\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n}(\pi\frac{\theta^2}{\theta+1}(1+y_i)\exp(-\theta y_i))^{x_i} + ((1-\pi)\lambda\exp(\lambda - y_i))^{1-x_i}, \quad (4.1)$$

with missing data $\boldsymbol{x}$. As parameters are independent the relevant loglikelihoods parts are

- $\ell(\theta|\boldsymbol{x}, \boldsymbol{y}) = 2\log(\theta)\sum_{i=1}^{n} x_i - \log(\theta+1)\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i y_i + c$
- $\ell(\lambda|\boldsymbol{x}, \boldsymbol{y}) = \log(\lambda)\sum_{i=1}^{n}(1-x_i) - \lambda\sum_{i=1}^{n}(1-x_i)y_i + c$

θ can be estimated as in section 1 replacing $\bar{y}$ with $(\sum_{i=1}^{n} x_i y_i)/\sum_{i=1}^{n} x_i$. The derivative of $\ell(\lambda|\boldsymbol{x}, \boldsymbol{y})$ is

$$\ell'(\lambda|\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\lambda}(n - \sum_{i=1}^{n} x_i) - \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i y_i. \quad (4.2)$$

This can be solved analytically

$$\hat{\lambda} = \frac{n - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i y_i} \quad (4.3)$$