SF2955 Computer Intensive Methods
HA2 - Group 22

Valeri Ivinskiy

June 5, 2020

Problem 1.

In this problem, we are going to analyse occurrence of British coal mine disasters with $d-1$ breakpoints. The data spans from 1851 to 1963. Let $t_1 = 1851$, $t_{d+1} = 1963$ and $t_i, i = 2, ..., d$ be the breakpoints. We collect these points in a vector $\boldsymbol{t} = (t_1, ..., t_{d+1})$. The disaster intensity in each interval is λ_i and we let $\boldsymbol{\lambda} = (\lambda_i, ..., \lambda_d)$. We denote the observed time points by $\boldsymbol{\tau} = (\tau_1, ..., \tau_n)$. We model the data on the interval $t_1 \leq t \leq t_{d+1}$ as:

$$\lambda(t) = \sum_{i=1}^{d} \lambda_i \mathbb{1}_{[t_i, t_{i+1}]}(t)$$

We compute the number of disasters in each sub-interval $t_1 \leq t \leq t_{d+1}$ as:

$$n_i(\boldsymbol{\tau}) = \sum_{j=1}^{n} \mathbb{1}_{[t_i, t_{i+1}]}(\tau_j)$$

A $\Gamma(2, \theta)$ prior is put on the intensities and a $\Gamma(2, \upsilon)$ hyperprior on θ, where υ is a fixed parameter which needs to be specified. We also put the following prior on the breakpoints:

$$f(\boldsymbol{t}) = \begin{cases} \prod_{i=1}^{d}(t_{i+1} - t_i), & \text{for } t_1 < t_2 < ... < t_d < t_{d+1} \\ 0, & \text{otherwise} \end{cases}$$

Then we have that

$$f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}) \propto \exp\left\{-\sum_{i=1}^{d}(\lambda_i(t_{i+1} - t_i)\right\} \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})}$$

We will proceed to sample from the posterior $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}|\boldsymbol{\tau})$ with a hybrid MCMC algorithm.

a.

To be able to construct a hybrid MCMC algorithm we have to compute the marginal posteriors.

$$\begin{aligned} f(\theta|\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) &\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}, \theta) f(\boldsymbol{\lambda}, \boldsymbol{t}, \theta) \\ &\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}) f(\boldsymbol{\lambda}|\theta) f(\theta) f(\boldsymbol{t}) \\ &\propto (\prod_{i=1}^{d} \lambda_i \exp\{-\lambda_i\theta\}\theta^2)(\theta \exp\{-\theta\upsilon\}) \\ &\propto \exp\left\{-(\upsilon + \sum_{i=1}^{d}\lambda_i)\theta\right\}\theta^{2d+1} \end{aligned} \tag{1}$$

where we see that this corresponds to a $\Gamma(2d+2, \upsilon + \sum_{i=1}^{d} \lambda_i)$-distribution.

$$
\begin{aligned}
f(\boldsymbol{\lambda}|\boldsymbol{t}, \boldsymbol{\tau}, \theta) &\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}, \theta) f(\boldsymbol{\lambda}, \boldsymbol{t}, \theta) \\
&\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}) f(\boldsymbol{\lambda}|\theta) f(\theta) f(\boldsymbol{t}) \\
&\propto \exp\left\{ -\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \right\} \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \prod_{i=1}^{d} \lambda_i \exp\{-\lambda_i \theta\} \qquad (2)
\end{aligned}
$$

Which leads to

$$
f(\lambda_i|\boldsymbol{\tau}, \boldsymbol{t}, \theta) \propto \exp\{-(t_{i+1} - t_i + \theta)\lambda_i\} \lambda_i^{n_i(\boldsymbol{\tau})+1} \qquad (3)
$$

which is a $\Gamma(n_i(\boldsymbol{\tau}) + 2, t_{i+1} - t_i + \theta)$ distribution. Lastly,

$$
\begin{aligned}
f(\boldsymbol{t}|\boldsymbol{\lambda}, \boldsymbol{\tau}, \theta) &\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}, \theta) f(\boldsymbol{\lambda}, \boldsymbol{t}, \theta) \\
&\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \boldsymbol{t}) f(\boldsymbol{\lambda}|\theta) f(\theta) f(\boldsymbol{t}) \\
&\propto \exp\left\{ -\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \right\} \prod_{i=1}^{d} (t_{i+1} - t_i) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \text{ for } t_1 < t_2 < ... < t_d < t_{d+1} \qquad (4)
\end{aligned}
$$

b.

We can now proceed with constructing a hybrid MCMC algorithm. For updating $\boldsymbol{t}$, we will choose the *random walk proposal*, as specified in the project description, with a pre-specified ρ. We also choose to do this for $N = 15000$ iterations. For the calculation of α, the acceptance ratio used is based on the full conditional of $\boldsymbol{t}_i$ given all other variables. Thus the algorithm would look like this:

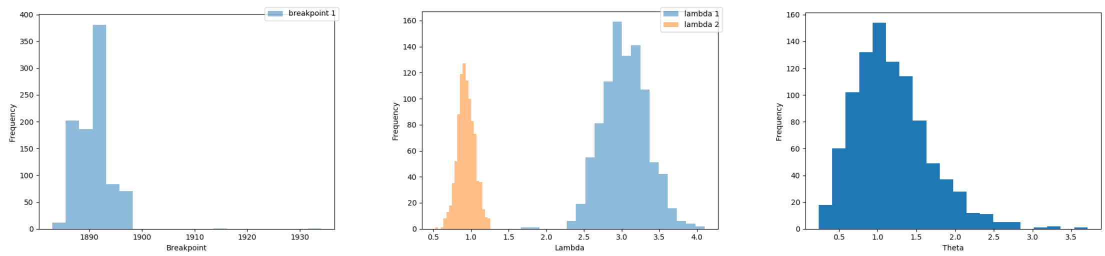
Algorithm 1 Hybrid MCMC algorithm

Initialize $\theta^0, \boldsymbol{\lambda}^0 = (\lambda_1^0, ..., \lambda_d^0), \boldsymbol{t}^0 = (t_1^0, ..., t_{d+1}^0)$
for $k = 0, ..., N-1$ **do**
    draw $\theta^{k+1} \sim \Gamma(2d+2, \upsilon + \sum_{i=1}^{d} \lambda_i^k)$
    for $i = 1, ..., d$ **do**
        draw $\lambda_i^{k+1} \sim \Gamma(n_i(\boldsymbol{\tau}) + 2, t_{i+1}^k - t_i^k + \theta^{k+1})$
    end for
    for $i = 2, ..., d$ **do**
        $R \leftarrow \rho(t_{i+1}^k - t_i^k)$
        $\epsilon \sim U(-R, R)$
        $t_i^* \leftarrow t_i^k + \epsilon$
        $\alpha \leftarrow \min(1, \frac{f(t_i^*|\boldsymbol{\lambda}^{k+1}, \boldsymbol{\tau}, \theta^{k+1})}{f(t_i^k|\boldsymbol{\lambda}^{k+1}, \boldsymbol{\tau}, \theta^{k+1})})$
        draw $u \sim U(0,1)$
        if $u \leq \alpha$ **then**
            $t_i^{k+1} \leftarrow t_i^*$
        else
            $t_i^{k+1} \leftarrow t_i^k$
        end if
    end for
end for

c.

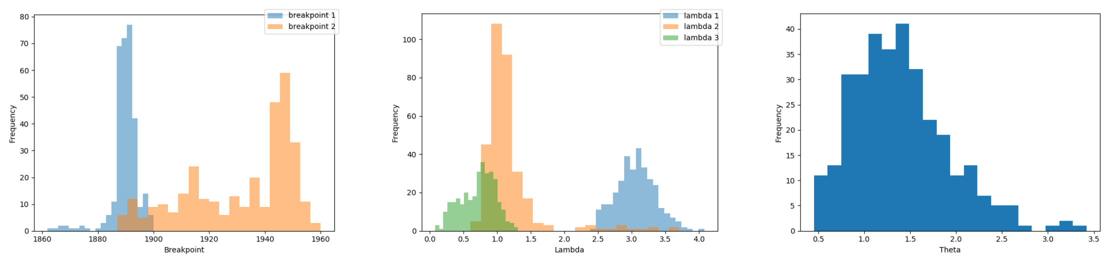
We chose to set $\upsilon = 1, \rho = 0.5$ and then we got the following plots for the breakpoints, λ and θ for each number of breakpoints:

Figure 1: 1 Breakpoint



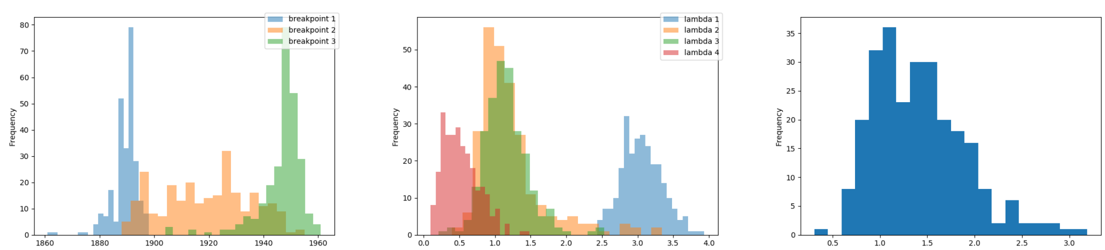
(a) Breakpoints (b) λ (c) θ

Figure 2: 2 Breakpoints



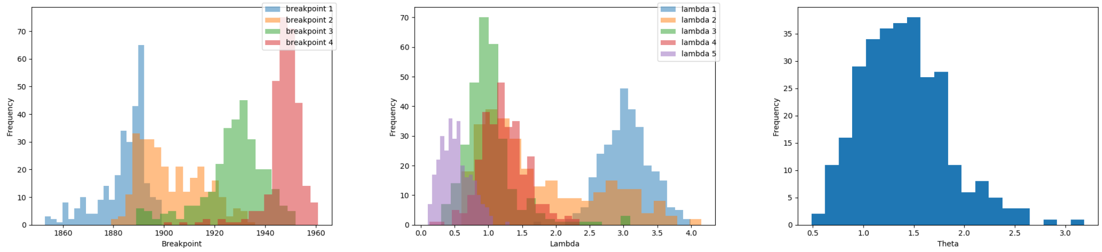
(a) Breakpoints (b) λ (c) θ

Figure 3: 3 Breakpoints



(a) Breakpoints (b) λ (c) θ

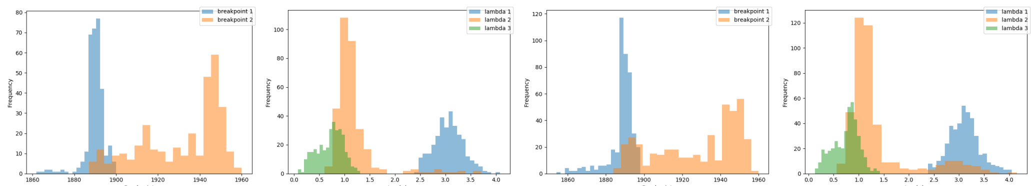
Figure 4: 4 Breakpoints



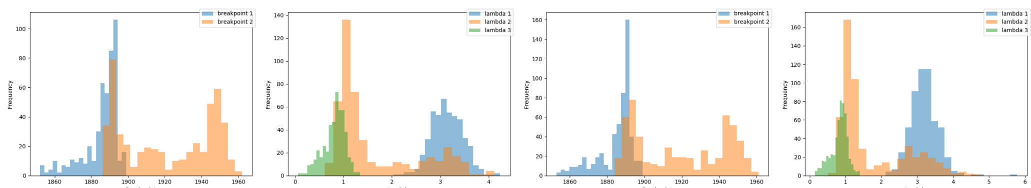
(a) Breakpoints (b) λ (c) θ

d.

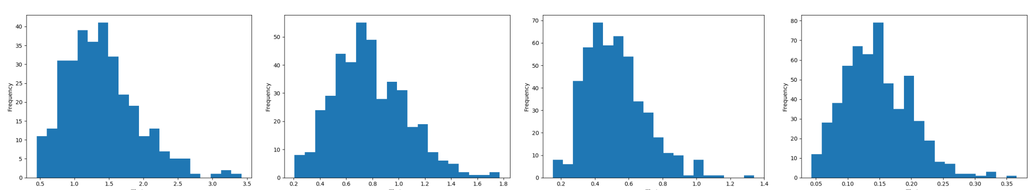
We must now investigate the sensitivity to the choice of the hyperparameter υ. To do this we specify $\rho = 0.5$ and $d = 2$, and run the algorithm over $\upsilon = [1, 5, 10, 50]$. This yields the following plots:



(a) t for $\upsilon = 1$ (b) λ for $\upsilon = 1$ (c) t for $\upsilon = 5$ (d) λ for $\upsilon = 5$



(a) t for $\upsilon = 10$ (b) λ for $\upsilon = 10$ (c) t for $\upsilon = 50$ (d) λ for $\upsilon = 50$

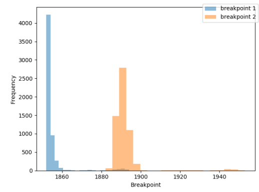


(a) θ for $\upsilon = 1$ (b) θ for $\upsilon = 5$ (c) θ for $\upsilon = 10$ (d) θ for $\upsilon = 50$

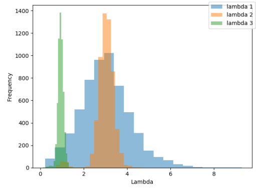
We see here that the breakpoints and intensities stay intact and insensitive for all different values of υ. The acceptance rates are also remaining intact, which we can determine by looking at the frequencies of each histogram. Regarding θ, the choice of υ seems to have a much larger impact, with the θ-values becoming smaller as υ increases. This is perhaps expected as the conditional distribution of θ given all other variables is heavily dependent on υ. Thus we can conclude that the choice of υ has an impact on θ.

e.

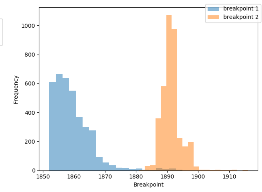
Now we will examine how sensitive the mixing and the posteriors are to the choice of ρ in the proposal distribution. We specify $\upsilon = 1$ and $d = 2$, and run the algorithm over $\rho = [0.05, 0.1, 0.5, 1]$. This yields the following plots:



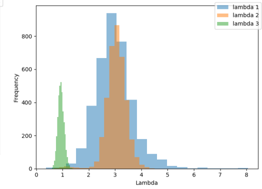
(a) Breakpoints for $\rho = 0.05$



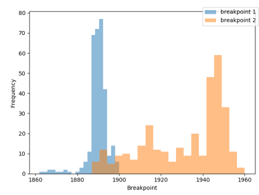
(b) λ for $\rho = 0.05$



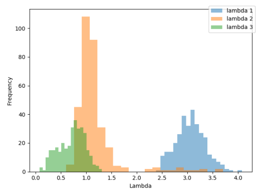
(c) Breakpoints for $\rho = 0.1$



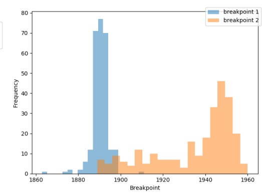
(d) λ for $\rho = 0.1$



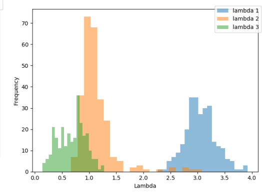
(a) Breakpoints for $\rho = 0.5$



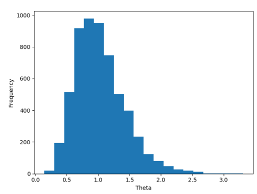
(b) λ for $\rho = 0.5$



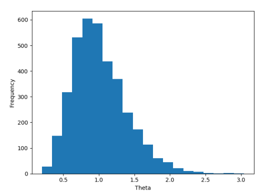
(c) Breakpoints for $\rho = 1$



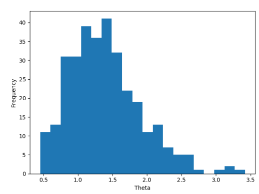
(d) λ for $\rho = 1$



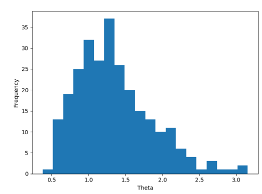
(a) θ for $\rho = 0.05$



(b) θ for $\rho = 0.1$



(c) θ for $\rho = 0.5$



(d) θ for $\rho = 1$

We see that the choice of ρ has a very large impact on the mixing, or the acceptance rate, by looking at the different plots of the breakpoints. We see that the frequencies are decreased (thus acceptance rate goes down) by several order of magnitudes as ρ increases. The λ's seem somewhat affected as well, but they remain within the same order of magnitude and it is rather their distribution which is changed as the acceptance rate goes down, and it remains more or less the same for $\rho = 0.05, 0.1$ and $\rho = 0.5, 1$ thus it is possible to conclude that the choice of ρ does not affect λ. We also see that the θ's are insensitive to the choice of ρ.

Problem 2.

A mixture model comprises an unobservable {0,1}-valued random variable X (referred to as index) such that $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = \theta$ and an observable random variable Y such that:

$$Y|X = 0 \sim g_0(y)dy$$
$$Y|X = 1 \sim g_1(y)dy$$

where g_0, g_1 are known probability densities. In this case the probability densities are Gaussian with means 0 and 1 respectively and the standard deviations 1 and 2 respectively. The probability of θ is unknown. We are given observations $\boldsymbol{y} = (y_1, ..., y_n)$ and denote by $\boldsymbol{x} = (x_1, ..., x_n)$ the corresponding unobserved index variables. Our aim is to compute a maximum likelihood estimator for θ using the EM algorithm.

a.

We will begin with stating the complete data log-likelihood function $\theta \mapsto \log f_\theta(\boldsymbol{x}, \boldsymbol{y})$.

We know that $f_\theta(\boldsymbol{x}, \boldsymbol{y}) = f_\theta(\boldsymbol{y}|\boldsymbol{x}) f_\theta(\boldsymbol{x})$ where both the factors are already given in the problem description. The likelihood function is thus given by:

$$f_\theta(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n} (g_0(y_i)(1-\theta)(1-x_i) + g_1(y_i)\theta x_i) \tag{5}$$

which yields the log-likelihood function:

$$\log f_\theta(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \log(g_0(y_i)(1-\theta)(1-x_i) + g_1(y_i)\theta x_i) \tag{6}$$

b.

Now we want to compute the conditional distribution $f_\theta(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_\theta(\boldsymbol{x},\boldsymbol{y})}{f_\theta(\boldsymbol{y})}$. The denominator is given by:

$$\begin{aligned} f_\theta(y_i) &= f_\theta(y_i|x_i = 1)p_\theta(x_i = 1) + f_\theta(y_i|x_i = 0)p_\theta(x_i = 0) \\ &= g_0(y_i)(1-\theta) + g_1(y_i)\theta \end{aligned}$$

And the conditional distribution is given by:

$$f_\theta(\boldsymbol{x}|\boldsymbol{y}) = \prod_{i=1}^{n} \frac{g_0(y_i)(1-\theta)(1-x_i) + g_1(y_i)\theta x_i}{g_0(y_i)(1-\theta) + g_1(y_i)\theta} \tag{7}$$

c.

We begin by inspecting the data in a histogram, as seen below.

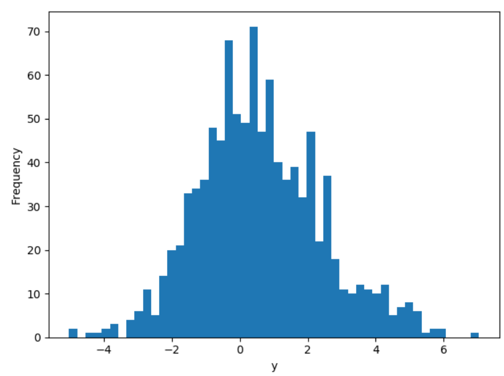


Figure 11: Histogram of mixture observations

And for the EM algorithm we begin by computing

$$\begin{aligned} Q_{\theta_\ell}(\theta) &= E_{\theta_\ell}(\log f_\theta(\boldsymbol{x}, \boldsymbol{y})|\boldsymbol{y}) \\ &= E_{\theta_\ell}(\sum_{i=1}^{n} \log f_\theta(x_i, y_i)|y_i) \\ &= E_{\theta_\ell}(\log(g_0(y_i)(1-\theta)(1-x_i) + g_1(y_i)\theta x_i)) \\ &= \sum_{i=1}^{n} (\log(g_0(y_i)(1-\theta))\mathbb{P}_{\theta_\ell}(x_i = 0|y_i) + \log(g_1(y_i)\theta)\mathbb{P}_{\theta_\ell}(x_i = 1|y_i)) \\ &= \sum_{i=1}^{n} \frac{g_0(y_i)(1-\theta_\ell)\log(1-\theta) + g_1(y_i)\theta_\ell \log(\theta)}{g_0(y_i)(1-\theta_\ell) + g_1(y_i)\theta_\ell} \end{aligned} \tag{8}$$

To obtain argmax $Q_{\theta_\ell}(\theta)$ we must derive (8) w.r.t. θ and set the derivative to 0. This yields:

$$\begin{aligned}
Q'_{\theta_\ell}(\theta) &= \sum_{i=1}^{n}\Big(-\frac{g_0(y_i)(1-\theta_\ell)}{(1-\theta)(g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell)} \\
&\quad + \frac{g_1(y_i)\theta_\ell}{\theta(g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell)}\Big) = 0 \\
&\iff \sum_{i=1}^{n}\frac{g_0(y_i)(1-\theta_\ell)}{(1-\theta)(g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell)} = \sum_{i=1}^{n}\frac{g_1(y_i)\theta_\ell}{\theta(g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell)}
\end{aligned} \tag{9}$$

(9) gives that

$$\theta_{\ell+1} = \text{argmax } Q_{\theta_\ell}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{g_1(y_i)\theta_\ell}{g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell} \tag{10}$$

This result will be used in the EM algorithm below:

Algorithm 2 EM algorithm

1: Initialize θ_0
2: **for** $\ell = 0, ..., N-1$ **do**
3: $\quad \theta_{\ell+1} = \frac{1}{n}\sum_{i=1}^{n}\frac{g_1(y_i)\theta_\ell}{g_0(y_i)(1-\theta_\ell)+g_1(y_i)\theta_\ell}$
4: **end for**

Using $N = 100$ and an initial guess of $\theta_0 = 0.5$ and implementing the algorithm above yields the following curve:

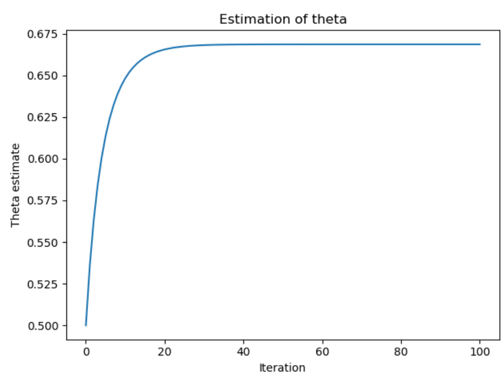


Figure 12: Estimation of θ

Here we also see that the final estimator is at $\theta = 0.6687$.