SF2955 Homework Assignment 2

Authors: Oliver Welin Odeback, Ludvig Eidmann

May 14, 2021

Bayesian analysis of coal mine disasters - constructing a complex MCMC algorithm

Problem 1

In this part we derive the marginal posteriors of $f(\theta|\lambda, t, \tau)$, $f(\lambda|\theta, t, \tau)$ and $f(t|\theta, t, \tau)$ respectively. For $f(\theta|\lambda, tt, \tau)$, we have that

$$f(\theta \mid \lambda, t, \tau) \propto f(\theta, \lambda, t, \tau) \propto f(\tau, t \mid \theta, \lambda) f(\lambda \mid \theta) f(\theta) \propto f(\tau \mid t, \theta, \lambda) f(t \mid \theta, \lambda) f(\lambda, \theta) f(\theta)$$

$$\propto f(\tau \mid t, \lambda) f(t) f(\lambda \mid \theta) f(\theta) \propto f\lambda \mid \theta) f(\theta) \propto (\prod_{i=1}^{d} \theta^2 \mathrm{e}^{-\theta\lambda_i}) \theta \mathrm{e}^{-\vartheta\theta} = \theta^{2d+1} \mathrm{e}^{-\theta(\vartheta + \sum_{i=1}^{d} \lambda_i)}$$

$$\propto \Gamma\left(\theta; 2(d+1), \vartheta + \sum_{i=1}^{d} \lambda_i\right)$$

For $f(\lambda|\theta, t, \tau)$, we have

$$f(\lambda|\theta, t, \tau) \propto f(\lambda, \tau, t, \theta) \propto f(\tau \mid t\lambda) f(\lambda \mid \theta) f(\theta) f(t)$$

$$\propto \exp\left\{-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i)\right\} \prod_{i=1}^{d} \lambda_i \prod \theta^2 \lambda_i \mathrm{e}^{-\theta\lambda_i} \cdot \vartheta^2 \theta \mathrm{e}^{-\vartheta\theta} \cdot f(t)$$

$$= \prod_{i=1}^{d} \mathrm{e}^{-\lambda_i (t_{i+1} - t_i)} \lambda_i^{n_i(\tau)+1} \mathrm{e}^{-\theta\lambda_i} = \lambda_i^{n_i(\tau)+1} \mathrm{e}^{-\lambda(\theta + t_{i+1} - t_i)}$$

$$\propto \Gamma\left(\lambda_i; n_i(\tau) + 2, \theta + t_{i+1} - t_i\right)$$

and hence, we have for $f(t|\theta, t, \tau)$ that

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

We continue by implementing a hybrid MCMC algorithm using a random walk proposal in order to update t. With parameter settings of $\vartheta = 4$, $\rho = 0.01$, $m = 100,000$, with a 1000 burn-in, we the expected values for breakpoints 1,...,4 (d=2,...,5) displayed in Table 1.

1 breakpoint	2 breakpoints	3 breakpoints	4 breakpoints
1891	1889	1889	1881
	1931	1923	1894
		1939	1925
			1944

Table 1: Expected values for different number of breakpoints with

Next, we investigate the sensitivity of the posteriors to the hyperparameters ϑ and ρ.

For a fixed value of ρ, set to $\rho = 10$, we implement the algorithm for d $=$ 2 and for different values of ϑ. Figure 1 and 2 shows the histogram of the distribution of the parameters for $\vartheta = 0.1$, $\vartheta = 5.05$ and $\vartheta = 10$. We see that of all the parameters, only θ vary with different values of ϑ.

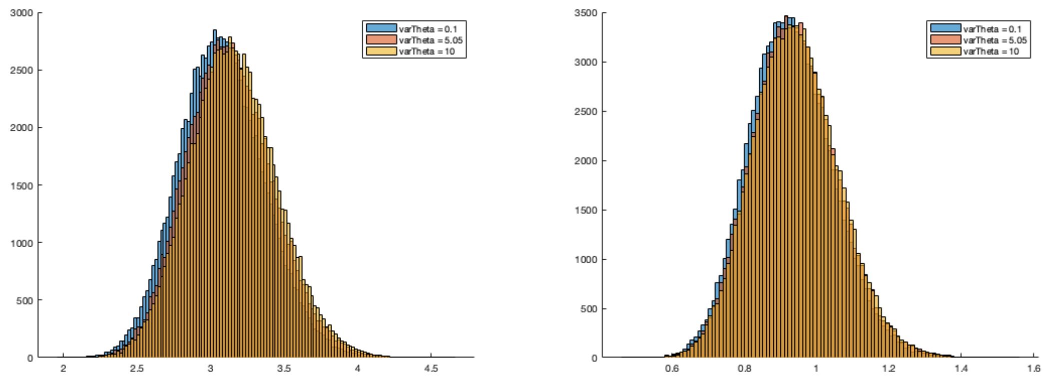

Figure 1: Histogram of distribution of λ_1 (left) and λ_2 (right)for different values of ϑ.

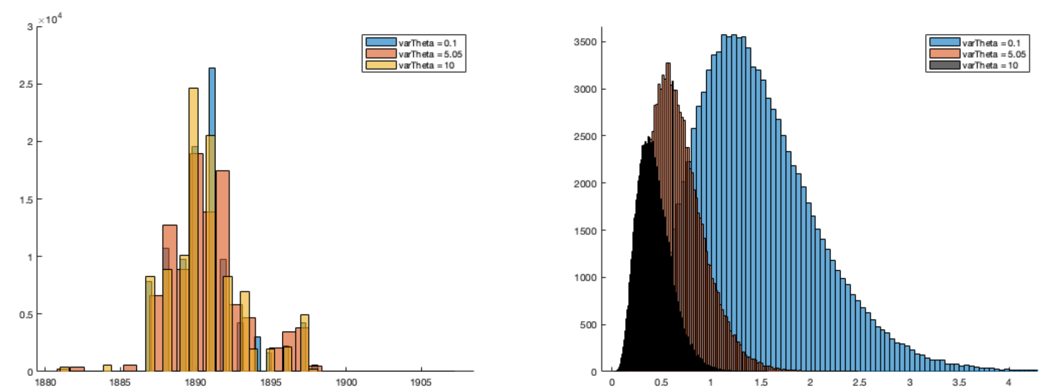

Figure 2: Histogram of distribution of t_1 (left) and θ (right) for different values of ϑ.

Finally, we investigate the sensitivity of the posteriors to the choice of ρ. We fix $\vartheta = 5$, and try for $\rho = 0.1$, $\rho = 5.05$ and $\rho = 10$. The implementation is displayed in Figure 3 and 4. None of the parameters seems to depend on the value chosen.

Furthermore, Figure 5-7 displays the MCMC trajectories associated with three different parameters $\rho = 0.1$, $\rho = 5.05$ and $\rho = 10$. along with the estimated autocorrelation functions. The mixing investigation (using autocorrelation function) reveals that the appropriate value for ρ is in the order of magnitude of 0.1.

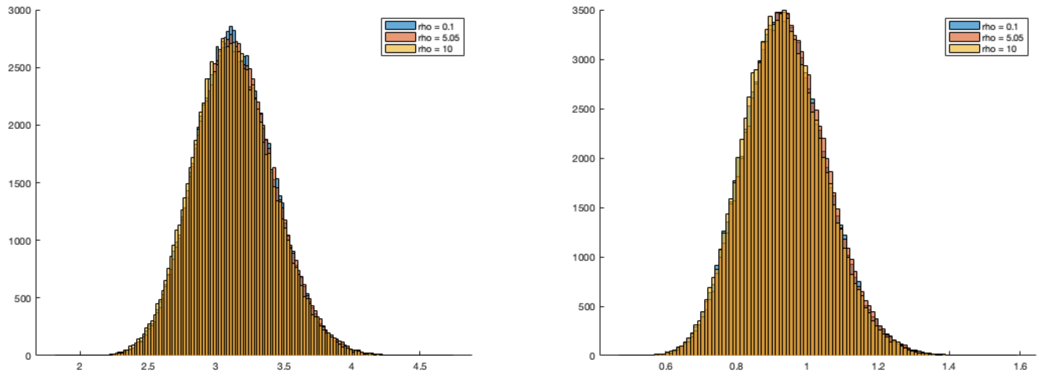

Figure 3: Histogram of distribution of λ_1 (left) and λ_2 (right) for different values of ρ.

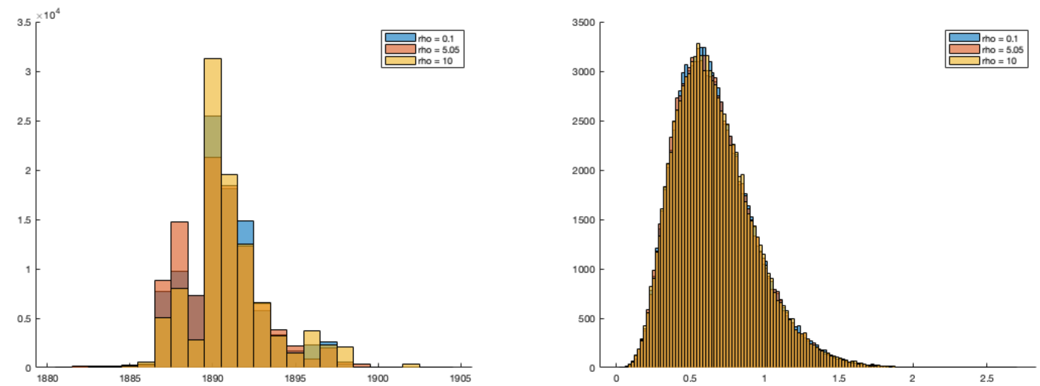

Figure 4: Histogram of distribution of t (left) and θ (right) for different values of ρ.

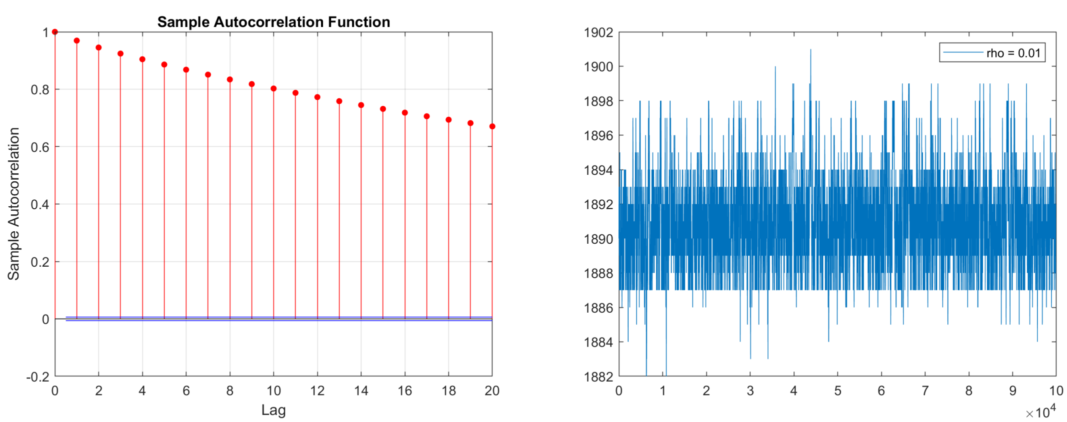

Figure 5: Autocorrelation function and MCMC trajectories for $\rho = 0.01$.

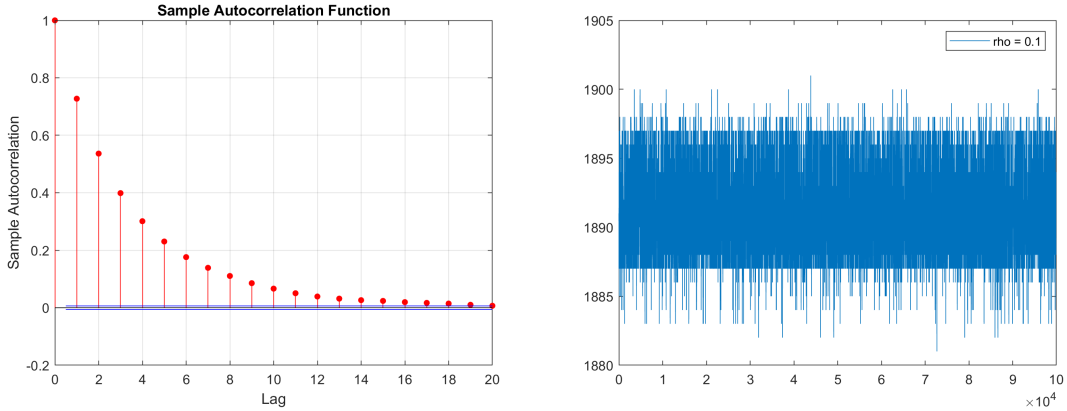

Figure 6: Autocorrelation function and MCMC trajectories for $\rho = 0.1$.

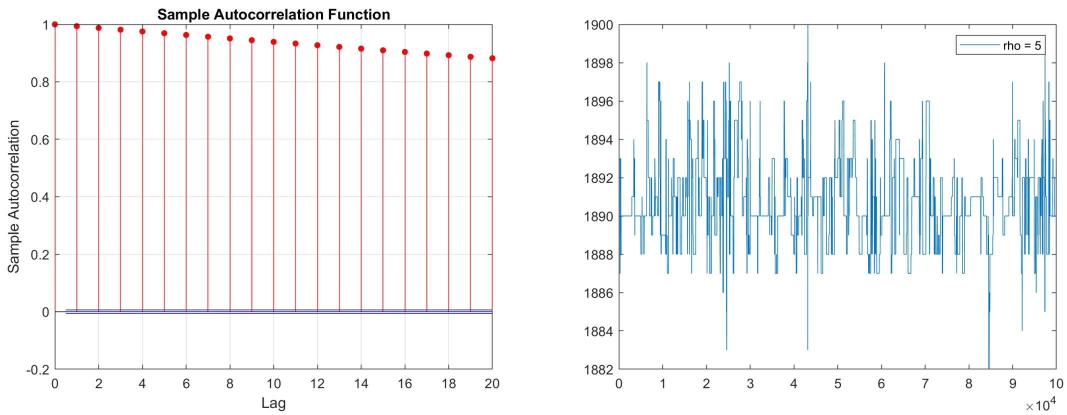

Figure 7: Autocorrelation function and MCMC trajectories for $\rho = 5$.

EM-based inference in mixture models

Problem 2 a) and b)

We begin by evaluating the expression for the complete data log-likelihood function $\theta \mapsto \log f_\theta(\mathbf{x} \mid \mathbf{y})$. Ignoring additive constants and using the fact that $\mu(x) = x, \sigma(x) = x + 1$, we have that

$$
\begin{aligned}
f_\theta(\mathbf{x}, \mathbf{y}) \propto f_\theta(\mathbf{y} \mid \mathbf{x}) f_\theta(\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2(x_i)}} \exp\left(-\frac{1}{2}\left(\frac{y_i - \mu(x_i)}{\sigma(x_i)}\right)^2\right) \theta^x (1-\theta)^{1-x} \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(x_i+1)^2}} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i}{x_i + 1}\right)^2\right) \theta^x (1-\theta)^{1-x} \\
&\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(x_i+1)^2}} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i}{x_i + 1}\right)^2 + x_i \log\left(\frac{\theta}{1-\theta}\right)\right)
\end{aligned}
$$

Now, taking a look at the logarithm and ignoring additive constants, this yields

$$
\log f_\theta(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} -\log(x_i + 1) - \frac{(y_i - x_i)^2}{2(x_i+1)^2} + \log\left(\frac{\theta}{1-\theta}\right) x_i
$$

which concludes the derivation of the data log-likelihood function.

Next, we evaluate the conditional distribution $f_\theta(\mathbf{x} \mid \mathbf{y})$. The distribution is given by

$$
\begin{aligned}
f_\theta(\mathbf{x} \mid \mathbf{y}) = \frac{f_\theta(\mathbf{x}, \mathbf{y})}{f_\theta(\mathbf{y})} \propto f_\theta(\mathbf{x}, \mathbf{y}) &\propto \prod_{i=1}^{n} \frac{\exp\left\{-\frac{(y_i - x_i)^2}{2(x_i+1)^2}\right\}}{x_i + 1} \theta^{x_i} (1-\theta)^{1-x_i} \\
&\propto \prod_{i=1}^{n} \frac{\exp\left\{-\frac{(y_i - x_i)^2}{2(x_i+1)^2} + x_i \log\left(\frac{\theta}{1-\theta}\right)\right\}}{x_i + 1}
\end{aligned}
$$

After normalizing we get

$$
f_\theta(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{n} \frac{\exp\left\{-\frac{(y_i - x_i)^2}{2(x_i+1)^2}\right\} \theta^{x_i} (1-\theta)^{1-x_i} / (x_i + 1)}{\mathrm{e}^{-y_i^2/2}(1-\theta) + \frac{\mathrm{e}^{-(y_i-1)^2/8}}{2}\theta}
$$

Problem 2 c)

In this section we implement the EM-algorithm in order to estimate the value of θ. A histogram of the observations contained in the file `mixture-observations.csv` is displayed in Figure 8.

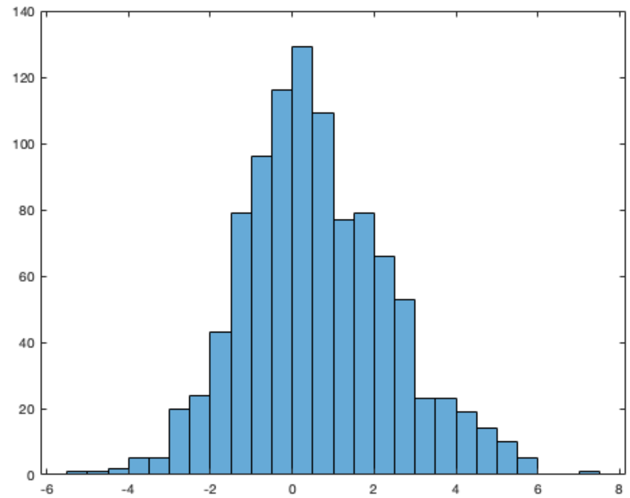

Figure 8: Histogram of $n = 1000$ observations.

Next, we derive the EM-updating formula for θ. Q_{θ_l} is given by

$$Q_{\theta_l}(\theta) = \sum_{i=1}^{n} \mathbb{E}_{\theta_l}[-\log(x_i+1) - \frac{(y_i-1)^2}{2(x_i+1)^2} + x_i \log\theta + (1-x_i)\log(1-\theta)]$$

$$= \sum_{i=1} f_{\theta_l}(0|y_i)(\frac{y_i^2}{2} + \log(1-\theta)) + f_{\theta_l}(1|y_i)(-\log 2 - \frac{(y_i-1)^2}{8} + \log(\theta))$$

$$\frac{\partial Q_{\theta_l}}{\partial\theta} = f_{\theta_l}(0|y_i)(\frac{-1}{1-\theta}) + f_{\theta_l}(1|y_i)/\theta = 0$$

$$\implies \theta = \frac{1}{n}(\sum_{i=1}^{n} f_{\theta_l}(1|y_i)).$$

We implement the EM algorithm for the initial value of $\theta_1 = 0.5$. For large N, this yields the final estimate $\hat{\theta} = 0.669$. The EM learning curve is displayed in Figure 9.

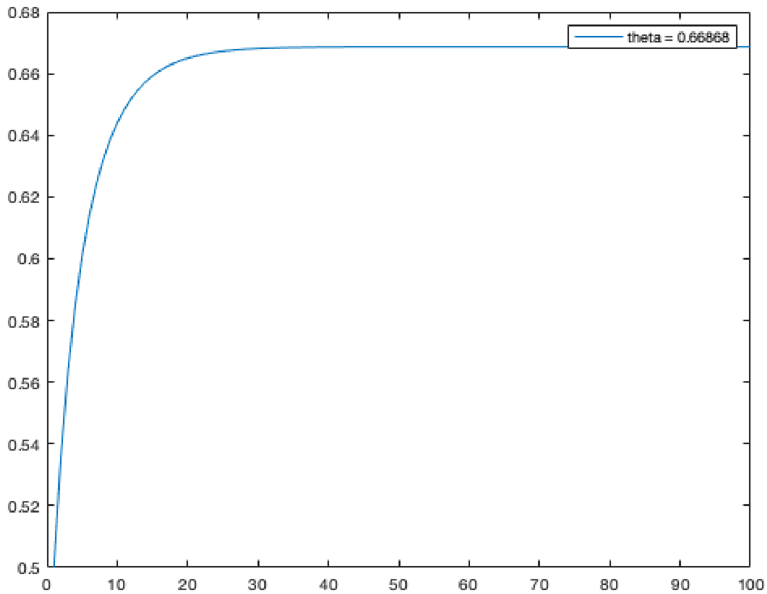

Figure 9: The EM learning curve with initial value $\theta_1 = 0.5$.