Consider the following hierarchical changepoint model for the number of occurrences Y_i of some event during time interval i with change point k.

$$\begin{aligned} Y_i|k,\theta,\lambda &\sim P(\theta) \text{ for } i=1,\dots,k \\ Y_i|k,\theta,\lambda &\sim P(\lambda) \text{ for } i=k+1,\dots,n \end{aligned}$$

Assume the following prior distributions:

$$\begin{aligned} \theta|b_1 \sim G(0.5,b_1), &\qquad \lambda|b_2 \sim G(0.5,b_2) \\ b_1 \sim G(1,1), &\qquad b_2 \sim G(1,1) \\ k \sim \text{Discrete Uniform}(1,\dots,n), & \end{aligned}$$

where k,θ,λ are independent and b_1,b_2 are independent.

Inference about this model is therefore based on the 5-dimensional posterior distribution $f(k,\theta,\lambda,b_1,b_2|\mathbf{Y})$ where $\mathbf{Y}=(Y_1,\dots,Y_n)$. Apply this model to an old data set of British coal mining disasters (Carlin and Louis 2000).

1. Use the Metropolis-Hastings algorithm for simulating from the posterior distribution $f(k,\theta,\lambda,b_1,b_2|\mathbf{Y})$. Include an estimated density plot based on your samples for each of the 5 parameters.
2. The parameters of interest here are k,θ,λ: report your MCMC estimates of their means.
3. Compute Monte Carlo standard errors for your mean estimates for k,θ,λ using the batch means method.

The kernel of the full joint (posterior) distribution is given by:

$$\pi(k,\theta,\lambda,b_1,b_2|\mathbf{Y}) \propto \left[\prod_{i=1}^{k} f(Y_i|\theta,\lambda,k)\right] g(\theta|b_1)h(b_1) \left[\prod_{i=k+1}^{n} f(Y_i|\theta,\lambda,k)\right] g(\lambda|b_2)h(b_2), \quad (1)$$

where f, g, and h denote the pdf's.

Based on the kernel above, we can find the form of the full conditional distributions for each parameter.

$$\pi(k|\mathbf{Y}) \propto \prod_{i=1}^{k} f(Y_i|\theta,\lambda,k) \prod_{i=k+1}^{n} f(Y_i|\theta,\lambda,k) \propto e^{-k(\theta-\lambda)}\theta^{\sum_{i=1}^{k} Y_i}\lambda^{\sum_{i=k+1}^{n} Y_i} = h(k|\mathbf{Y}). \quad (2)$$

$$\pi(\theta|k, b_1, \mathbf{Y}) \propto \left[\prod_{i=1}^{k} f(Y_i|\theta, \lambda, k)\right] g(\theta|b_1) \propto \theta^{\sum_{i=1}^{k} Y_i+1/2-1} e^{-\theta(k+1/b_1)}$$

$$\sim Gamma\left(\sum_{i=1}^{k} Y_i + 0.5, \frac{1}{k+1/b_1}\right). \tag{3}$$

Similarly, we have:

$$\pi(\lambda|k, b_2, \mathbf{Y}) \sim Gamma\left(\sum_{i=k+1}^{n} Y_i + 0.5, \frac{1}{n-k+1/b_2}\right). \tag{4}$$

$$\pi(b_1|\theta) \propto g(\theta|b_1)h(b_1) \propto b_1^{-1/2} e^{-\theta/b_1 - b_1} = h(b_1|\theta), \text{ and} \tag{5}$$

$$\pi(b_2|\lambda) \propto g(\lambda|b_2)h(b_2) \propto b_2^{-1/2} e^{-\lambda/b_2 - b_2} = h(b_2|\lambda). \tag{6}$$

The Block-at-a-time Metropolis-Hastings algorithm is used for simulating from the posterior distribution $f(k, \theta, \lambda, b_1, b_2|\mathbf{Y})$. Since the full conditions are known for θ and λ, we use Gibbs algorithm for updating theses two parameters. The proposal used for updating k is the Discrete Uniform $(1, n)$. For b_1 and b_2 the proposal transition kernel is *Gamma* density with parameter of shape (α) equal to the current value of b_1 and b_2, and $\beta = 1$. The algorithm below describes how to update the Markov chain in one step, one parameter at a time.

Algorithm:

Step 1. Choose initial values for $k^0, \theta^0, \lambda^0, b_1^0, b_2^0$.

Step 2. Generate $k^1 \sim DiscreteUniform(1, n)$ and accept k^1 with probability
$\alpha(k^0, k^1) = \min\left[1, \frac{h(k^1|\mathbf{Y})}{h(k^0|\mathbf{Y})}\right]$, where $h(\cdot)$ is given by (2).

Step 3. Generate $\theta^1 \sim Gamma\left(\sum_{i=1}^{k^1} Y_i + 0.5, \frac{1}{k^1 + 1/b_1^0}\right)$.

Step 4. Generate $\lambda^1 \sim Gamma\left(\sum_{i=k^1+1}^{n} Y_i + 0.5, \frac{1}{n - k^1 + 1/b_2^0}\right)$.

Step 5. Generate $b_1^1 \sim Gamma(b_1^0, 1)$ and accept b_1^1 with probability
$\alpha(b_1^0, b_1^1) = \min\left[1, \frac{h(b_1^1|\theta^1)}{h(b_1^0|\theta^1)} \frac{Gamma(b_1^0; b_1^1, 1)}{Gamma(b_1^1; b_1^0, 1)}\right]$,
where $h(\cdot|\theta^1)$ is given by (5).

Step 6. Generate $b_2^1 \sim Gamma(b_2^0, 1)$ and accept b_2^1 with probability

$$\alpha(b_2^0, b_2^1) = \min\left[1, \frac{h(b_2^1|\lambda^1)\, Gamma(b_2^0; b_2^1, 1)}{h(b_2^0|\lambda^1)\, Gamma(b_2^1; b_2^0, 1)}\right],$$

where $h(\cdot|\lambda^1)$ is given by (6).

Step 7. Repeat Steps 2-6 until desired number of updates.

The length of the Markov chain in this simulation was 102400. The initial values used were: $k^0 = 56, \theta^0 = \frac{1}{56}\sum_{i=1}^{56} Y_i, \lambda^0 = \frac{1}{56}\sum_{i=57}^{112} Y_i, b_1^0 = 1, b_2^0 = 1$. The estimated density plots based on the coal data set for each of the 5 parameters are in Figure 1. Figure 2 presents the behavior of the MCMC estimates for the mean of each parameter when we increase the sample size. We can heuristically check the convergence of the estimates. In this case, we can say that the MCMC estimates converge quickly. Figure 3 shows us the ACF's plots. Note that the autocorrelations for θ, and λ are very small.

The acceptance rate, the MCMC estimates and the MC Standard Error (calculated by the batch means method discussed in class) are in the table bellow. The MC Standard Error for the estimates are small.

	k	θ	λ	b_1	b_2
Acceptance Rate	0.0528	1	1	0.575	0.553
MCMC Estimate	40.00867	3.0884	0.9156	2.2603	1.4677
MC Std. Error	0.0425	0.0017	0.0007	0.0103	0.0085

Table 1: Acceptance Rates, MCMC Estimates and MC Standard Error

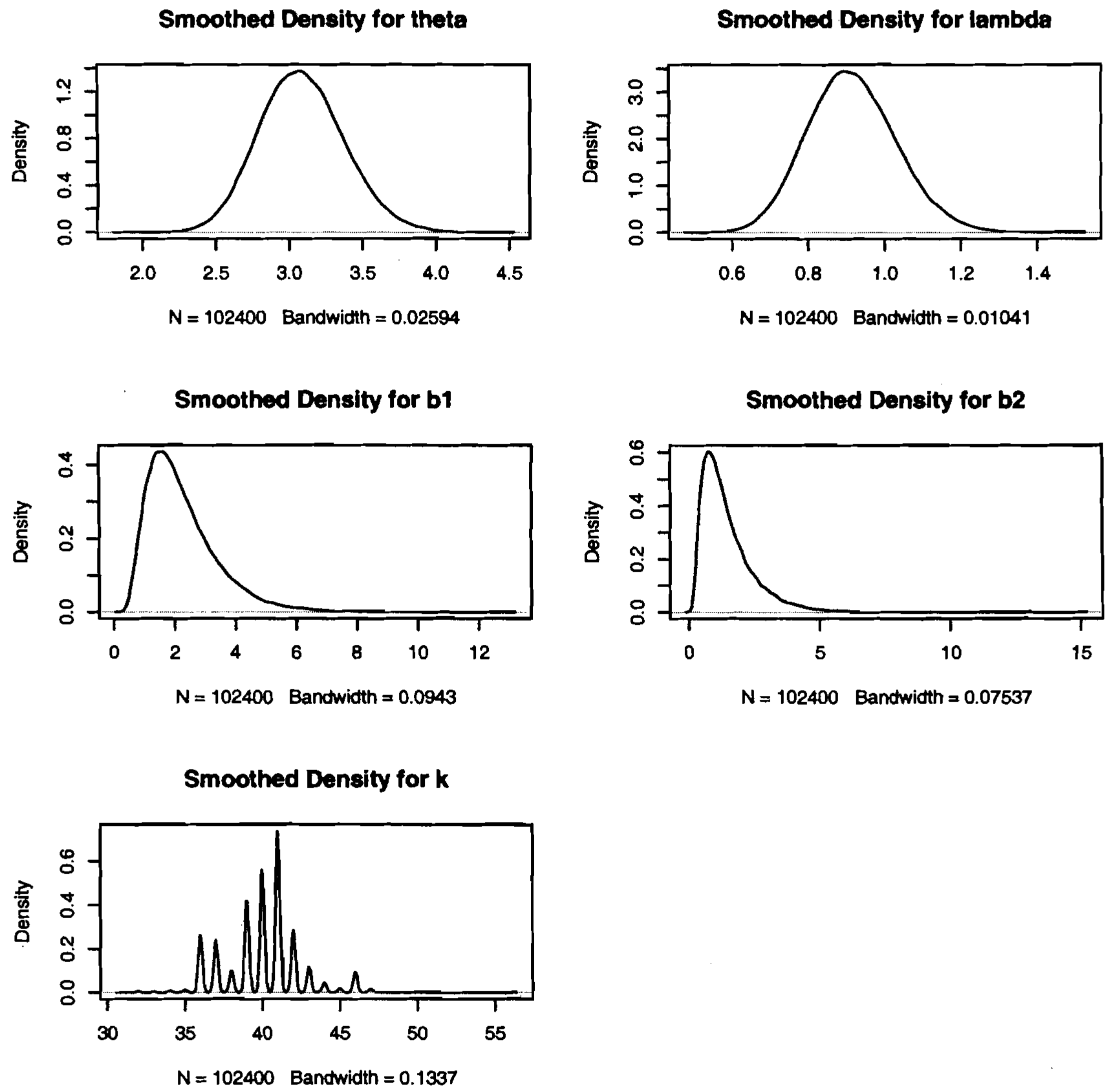

Figure 1: Smoothed Densities for $\theta, \lambda, b_1, b_2$, and k

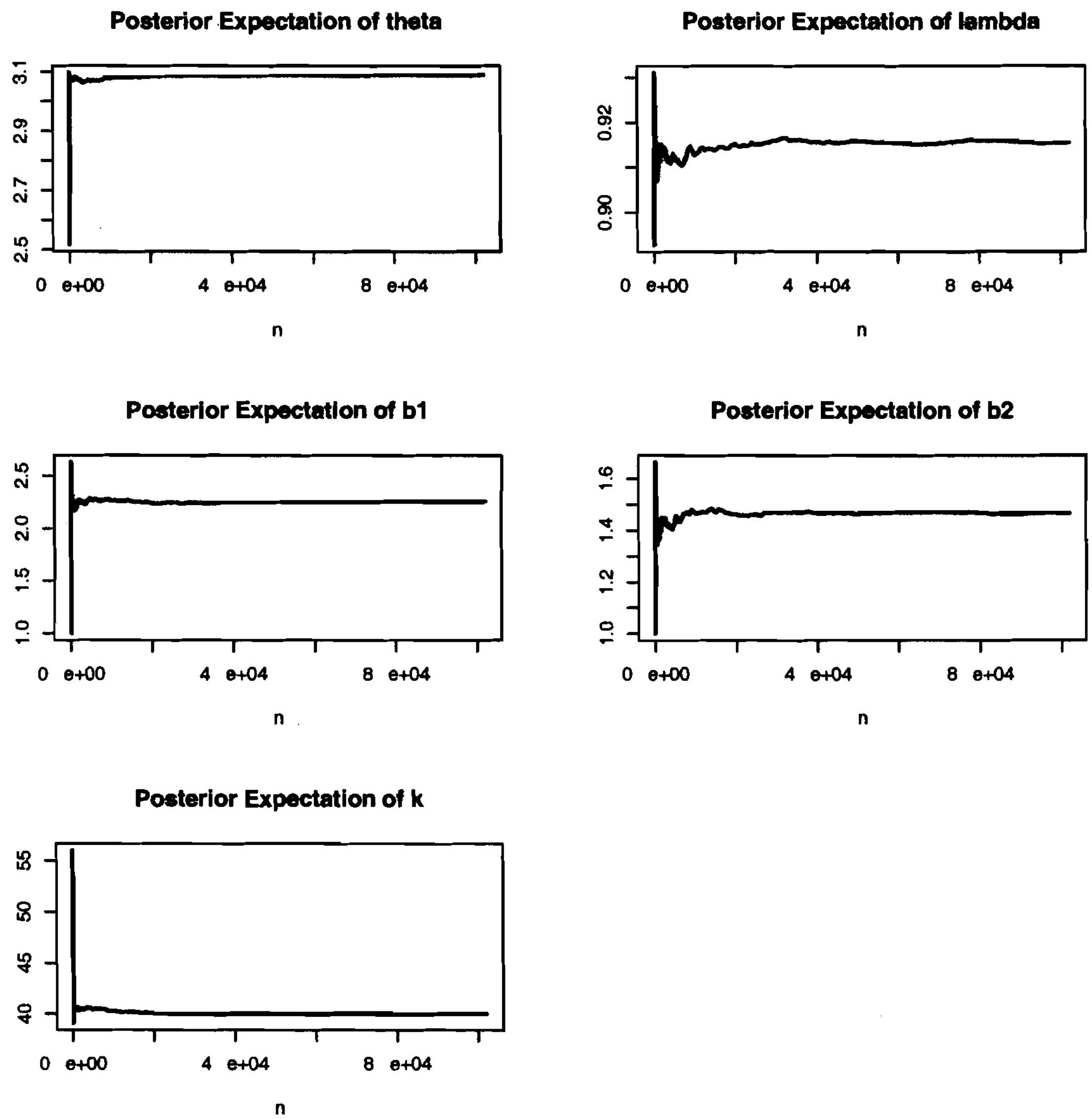

Figure 2: MCMC Estimates for $\theta, \lambda, b_1, b_2$, and k by n

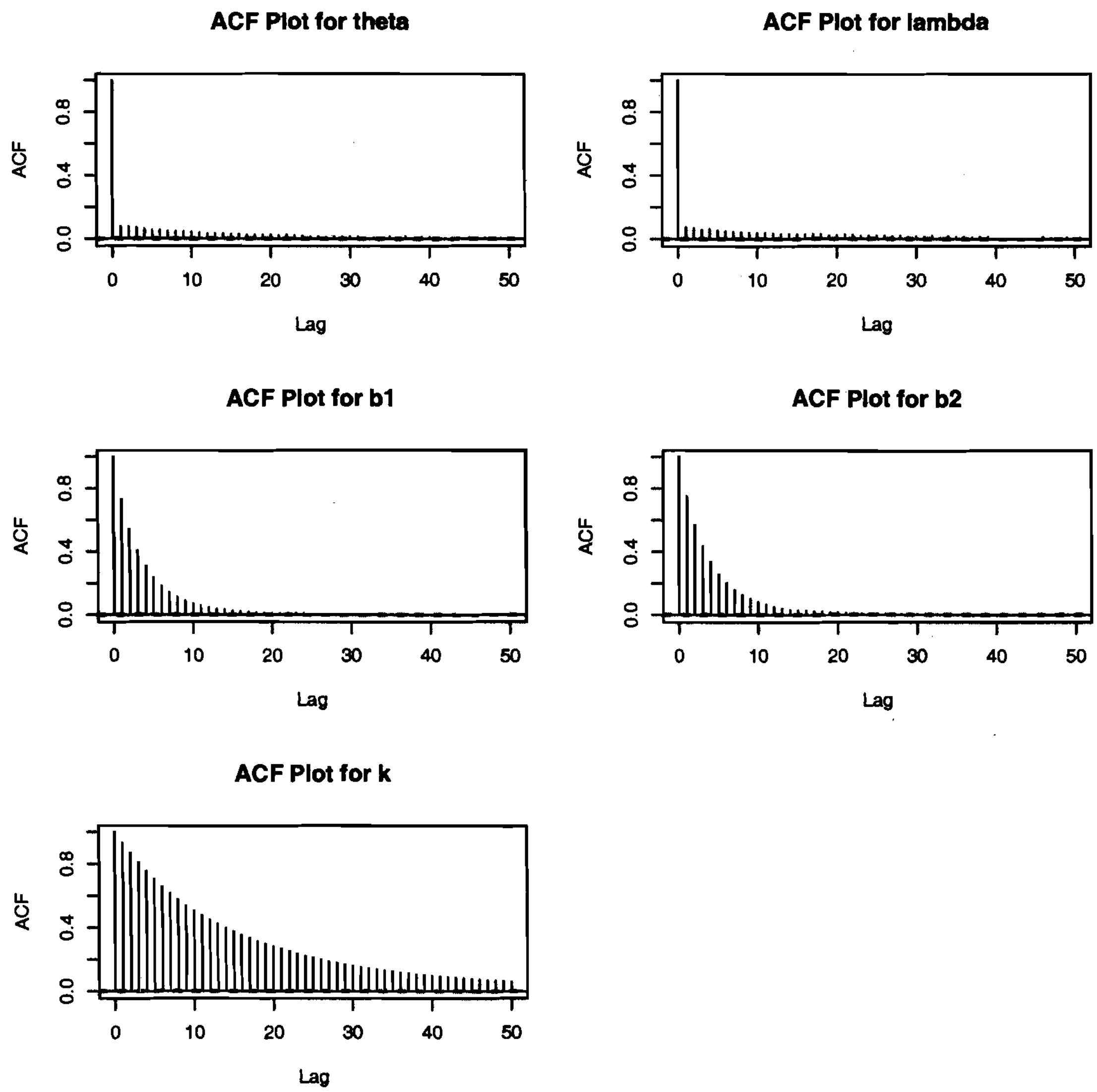

Figure 3: ACF Plots for $\theta, \lambda, b_1, b_2$, and k