Penn State Astrostatistics MCMC tutorial
Bayesian change point model with Gamma hyperpriors

The model is borrowed from Chapter 5 of "Bayes and Empirical Bayes Methods for Data Analysis" by Carlin and Louis (2000). This writeup was originally produced by Murali Haran of Penn State's Department of Statistics.

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

$$\begin{aligned} Y_i|k,\theta,\lambda \sim & \text{Poisson}(\theta) \text{ for } i=1,\ldots,k; \\ Y_i|k,\theta,\lambda \sim & \text{Poisson}(\lambda) \text{ for } i=k+1,\ldots,n. \end{aligned} \tag{1}$$

We assume that the Y_i are mutually independent conditional on the parameters.

Assume the following prior distributions on the parameters:

$$\begin{aligned} \theta|b_1 \sim & \text{Gamma}(0.5, b_1) \qquad (\text{pdf}=g_1(\theta|b_1)) \\ \lambda|b_2 \sim & \text{Gamma}(0.5, b_2) \qquad (\text{pdf}=g_2(\lambda|b_2)) \\ b_1 \sim & \text{IG}(0,1) \qquad (\text{pdf}=h_1(b_1)) \\ b_2 \sim & \text{IG}(0,1) \qquad (\text{pdf}=h_2(b_2)) \\ k \sim & \text{Uniform}(1,\ldots,n) \qquad (\text{pmf }=u(k)) \end{aligned} \tag{2}$$

Here, we assume k, θ, and λ are conditionally independent, and b_1, b_2 are independent. We use the gamma(α,β) and IG(α,β)—or inverse gamma(α,β)—density parameterizations

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \text{ and } \frac{e^{-1/\beta x}}{\Gamma(\alpha)\beta^\alpha x^{\alpha+1}},$$

respectively.

Inference for this model is based on the 5-dimensional **posterior** distribution $f(k,\theta,\lambda,b_1,b_2|\mathbf{Y})$, where $\mathbf{Y}=(Y_1,\ldots,Y_n)$. The posterior distribution is obtained up to a multiplicative constant by multiplying the likelihood (the density of $\mathbf{Y}$ given the parameters) times the joint prior of the parameters. This gives

$$\begin{aligned} f(k,\theta,\lambda,b_1,b_2|\mathbf{Y}) \;\propto\; & \prod_{i=1}^{k} f_1(Y_i|\theta,\lambda,k) \prod_{i=k+1}^{n} f_2(Y_i|\theta,\lambda,k) \\ & \times g_1(\theta|b_1) g_2(\lambda|b_2) h_1(b_1) h_2(b_2) u(k) \\ = & \prod_{i=1}^{k} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \prod_{i=k+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} \\ & \times \frac{1}{\Gamma(0.5) b_1^{0.5}} \theta^{-0.5} e^{-\theta/b_1} \times \frac{1}{\Gamma(0.5) b_2^{0.5}} \lambda^{-0.5} e^{-\lambda/b_2} \\ & \times \frac{1}{\Gamma(c_1) d_1^{c_1}} b_1^{c_1-1} e^{-b_1/d_1} \frac{1}{\Gamma(c_2) d_2^{c_2}} b_2^{c_2-1} e^{-b_2/d_2} \times \frac{1}{n}. \end{aligned} \tag{3}$$

If we are able to draw samples from this distribution, we can answer questions of interest.

Full Conditional Distributions of Parameters

Our goal is to draw samples from the 5-dimensional **posterior** distribution in Equation (3). The reason (3) gives a formula for what f is proportional to instead of what it equals exactly (hence $\propto$ rather than $=$) is because the missing multiplicative constant can only be computed by integrating the function. Fortunately, the Metropolis-Hastings algorithm does not require knowledge of this normalizing constant.

From (3) we can obtain full conditional distributions for each parameter by ignoring all terms that are constant with respect to the parameter. Sometimes these full conditional distributions are well known distributions such as gamma or normal:

- Full conditional for θ:

$$\begin{aligned} f(\theta|k,\lambda,b_1,b_2,\mathbf{Y}) &\propto \prod_{i=1}^{k} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \times \frac{1}{\Gamma(0.5) b_1^{0.5}} \theta^{-0.5} e^{-\theta/b_1} \\ &\propto \theta^{\sum_{i=1}^{k} Y_i - 0.5} e^{-\theta(k+1/b_1)} \\ &\propto \text{Gamma}\left(\sum_{i=1}^{k} Y_i + 0.5, \frac{b_1}{k b_1 + 1} \right). \end{aligned} \tag{4}$$

- Full conditional for λ:

$$\begin{aligned} f(\lambda|k,\theta,b_1,b_2,\mathbf{Y}) &\propto \prod_{i=k+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} \times \frac{1}{\Gamma(0.5) b_2^{0.5}} \lambda^{-0.5} e^{-\lambda/b_2} \\ &\propto \text{Gamma}\left(\sum_{i=k+1}^{n} Y_i + 0.5, \frac{b_2}{(n-k) b_2 + 1} \right). \end{aligned} \tag{5}$$

- Full conditional for k:

$$\begin{aligned} f(k|\theta,\lambda,b_1,b_2,\mathbf{Y}) &\propto \prod_{i=1}^{k} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \prod_{i=k+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} \\ &\propto \theta^{\sum_{i=1}^{k} Y_i} \lambda^{\sum_{i=k+1}^{n} Y_i} e^{-k\theta - (n-k)\lambda}. \end{aligned} \tag{6}$$

- Full conditional for b_1:

$$f(b_1|k,\theta,\lambda,b_2,\mathbf{Y}) \propto \frac{1}{b_1^{0.5}} e^{-\theta/b_1} \times \frac{e^{-1/b_1}}{b_1} \propto b_1^{-1.5} e^{-(1+\theta)/b_1} \propto IG(0.5, 1/(\theta+1)). \tag{7}$$

- Full conditional for b_2:

$$f(b_2|k,\theta,\lambda,b_1|\mathbf{Y}) \propto \times \frac{1}{b_2^{0.5}} e^{-\lambda/b_2} \times \frac{e^{-1/b_2}}{b_2} \propto b_2^{-1.5} e^{-(1+\lambda)/b_2} \propto IG(0.5, 1/(\lambda+1)). \tag{8}$$

Remark 1: The θ, λ, b_1, and b_2 parameters all have full conditional distributions that are well known and easy to sample from. We can therefore perform Gibbs updates on them where the draws are from their full conditionals. However, the full conditional for k is not a standard distribution, so we need to use the more general Metropolis-Hastings update instead of a Gibbs update.

Remark 2: The inverse gamma density is said to be a *conjugate* prior for b_1 and b_2 in this case since it results in a posterior that is also inverse gamma and therefore trivial to sample. As such, this density is mathematically convenient (due to its conjugacy property) but does not necessarily result in a better MCMC sampler. Also, it has poorly behaved moments; it may be better to adopt another prior density (such as a Gamma) instead.

The Metropolis-Hastings algorithm:

1. Pick a starting value for the Markov chain, say $(\theta^0, \lambda^0, k^0, b_1^0, b_2^0) = (1, 1, 20, 1, 1)$.
2. 'Update' each variable in turn at the ith iteration, $i = 1, \ldots, N$:
   (a) **Gibbs update of θ:** Sample θ^i from the gamma density of (4) using the most up-to-date values of k and b_1.
   (b) **Gibbs update of λ:** Sample λ^i from the gamma density of (5) using the most up-to-date values of k and b_2.
   (c) **Gibbs update of b_1:** Sample b_1^i from the inverse gamma density of (7) using the most up-to-date value of θ.
   (d) **Gibbs update of b_2:** Sample b_2^i from the inverse gamma density of (8) using the most up-to-date value of λ.
   (e) **Metropolis-Hastings update of k:** Sample $k \sim f(k|\theta, \lambda, b_1, b_2, \mathbf{Y})$ using the most up-to-date values of k, θ, λ, b_1, and b_2 as follows:
      i. Propose a new value for k, k^*, according to a proposal distribution, say, $q(k|\theta, \lambda, b_1, b_2, \mathbf{Y})$. In our example, we pick
      $$q(k|\theta, \lambda, b_1, b_2, \mathbf{Y}) = \text{Unif}\{2, \ldots, n-1\}.$$
      ii. Compute the Metropolis-Hastings accept-reject ratio
      $$\alpha(k, k^*) = \min\left\{\frac{f(k^*|\theta, \lambda, b_1, b_2, \mathbf{Y})q(k|\theta, \lambda, b_1, b_2, \mathbf{Y})}{f(k|\theta, \lambda, b_1, b_2, \mathbf{Y})q(k^*|\theta, \lambda, b_1, b_2, \mathbf{Y})}, 1\right\}.$$
      iii. With probability $\alpha(k, k^*)$, "accept" the value k^* (i.e., set $k^i = k^*$); otherwise, "reject" k^* (i.e., keep $k^i = k^{i-1}$).
   (f) You now have a new Markov chain state $(\theta^i, \lambda^i, k^i, b_1^i, b_2^i)$

A small modification

We'll now replace the priors of b_1 and b_2 in (2), as follows:

$$\begin{aligned} b_1 &\sim \text{Gamma}(c_1, d_1) \qquad (\text{pdf}=h_1(b_1)) \\ b_2 &\sim \text{Gamma}(c_2, d_2) \qquad (\text{pdf}=h_2(b_2)), \end{aligned}$$

where we take $c_1 = c_2 = 0.01$ and $d_1 = d_2 = 100$. It is possible to place prior distributions on the c_i and d_i hyperparameters, though here we simply assume that they are constant.

For the full conditionals, everything is the same for θ, λ, and k. However, we now get the following full conditionals for b_1 and b_2:

- for b_1:

$$f(b_1|k, \theta, \lambda, b_2, \mathbf{Y}) \propto \frac{1}{b_1^{0.5}} e^{-\theta/b_1} \times b_1^{c_1-1} e^{-b_1/d_1} \tag{9}$$

- for b_2:

$$f(b_2|k, \theta, \lambda, b_1|\mathbf{Y}) \propto \times \frac{1}{b_2^{0.5}} e^{-\lambda/b_2} \times b_2^{c_2-1} e^{-b_2/d_2} \tag{10}$$

$f(b_1|k, \theta, \lambda, b_2, \mathbf{Y})$ and $f(b_2|k, \theta, \lambda, b_1|\mathbf{Y})$ are not well-known densities. We can use a Metropolis-Hastings accept-reject step to sample from their full conditionals.