Penn State Astrostatistics MCMC tutorial
Murali Haran, Penn State Dept. of Statistics

Bayesian change point model: full conditional distributions

Our goal is to draw samples from the 5-dimensional **posterior** distribution $f(k, \theta, \lambda, b_1, b_2 | \mathbf{Y})$ The posterior distribution is

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

Note: The reason we have a formula for what f is proportional to (hence $\propto$ rather than $=$) instead of an exact description of the function is because the missing constant (the normalizing constant) can only be computed by integrating the above function. Fortunately, the Metropolis-Hastings algorithm does not require knowledge of this normalizing constant.

From (1) 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 the Gamma or Normal.

Full conditional for θ:

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

Full conditional for λ:

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

Full conditional for k:

$$f(k|\theta, \lambda, b_1, b_2, \mathbf{Y}) \propto \prod_{i=1}^{k} \frac{\theta^{Y_i} e^{-\theta}}{Y_i!} \prod_{i=k+1}^{n} \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!}$$

$$\propto \theta^{\sum_{i=1}^{k} Y_i} \lambda^{\sum_{i=k+1}^{n} Y_i} e^{-k\theta - (n-k)\lambda}.$$

Full conditional for b_1:

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

Full conditional for b_2:

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

We are now in a position to run the Metropolis-Hastings algorithm.

Note 1: $\theta, \lambda, b_1, b_2$ all have full conditional distributions that are well known and easy to sample from. We can therefore perform Gibbs updates on them where the draw is from their full conditional. 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.

Note 2: The Inverse Gamma density is said to be a **conjugate** prior 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:

   (a) Sample $\theta^i \sim f(\theta|k, \lambda, b_1, b_2, \mathbf{Y})$ using the most upto date values of k, λ, b_1, b_2 (Gibbs update using the derived Gamma density).

   (b) Sample $\lambda^i \sim f(\lambda|k, \theta, b_1, b_2, \mathbf{Y})$ using the most upto date values of k, θ, b_1, b_2. (Gibbs update using the derived Gamma density).

   (c) Sample $b_1^i \sim f(b_1|k, \theta, \lambda, b_2, \mathbf{Y})$ using the most upto date values of k, θ, λ, b_2. (Gibbs update using the derived Gamma density).

   (d) Sample $b_2^i \sim f(b_2|k, \theta, \lambda, b_1, \mathbf{Y})$ using the most upto date values of k, θ, λ, b_1. (Gibbs update using the derived Gamma density).

   (e) Sample $k \sim f(k|\theta, \lambda, b_1, b_2, \mathbf{Y})$ using the most upto date values of $k, \theta, \lambda, b_1, b_2$. This requires a Metropolis-Hastings update:

   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 simple example we pick $q(k|\theta, \lambda, b_1, b_2, \mathbf{Y}) = \mathrm{Unif}\{2, \ldots, m-1\}$ where m is the length of the vector (time series) $\mathbf{Y}$.

   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. Accept the new value k^* with probability $\alpha(k, k^*)$, otherwise 'reject' k^*, i.e., the next value of k remains the same as before.

   (f) You now have a new Markov chain state $(\theta^1, \lambda^1, k^1, b_1^1, b_2^1)$

3. Return to step #2 N-1 times to produce a Markov chain of length N.