Contributed Discussion on Article by Pratola*

Comment by Oksana A. Chkrebtii

Abstract. Pratola (2016) introduces a novel proposal mechanism for the Metropolis–Hastings step of a Markov chain Monte Carlo (MCMC) sampler that allows efficient traversal of the space of latent stochastic partitions defined by binary regression trees. Here we discuss two considerations: the first is the use of the new proposal mechanism within a population Markov chain Monte Carlo sampler (Geyer, 1991) to further increase sampling efficiency in the presence of greatly separated posterior modes, the second is a prior model that favors parsimony for the problem of variable selection.

Keywords: population Markov chain Monte Carlo, model selection, Bayesian treed regression.

We congratulate the author on an important contribution to sampling methodology for Bayesian treed regression. The joint posterior over the binary tree partition and model parameters is notoriously difficult to explore with existing local proposal mechanisms (Chipman et al., 2010; Wu et al., 2007; Gramacy and Lee, 2008). The rotation proposal step introduced by Pratola (2016) is important because it allows movement between disjoint high posterior density regions that arise when sampling regression tree structures. Given that this is a very challenging sampling problem, we suggest incorporating the proposal mechanism into a sampling scheme that can quickly move between posterior modes. We also note that improved ability to sample a possibly multimodal posterior allows us to consider the problem of variable selection, alluded to in the first motivating example in the paper where confounded variables tended to be added to the model together. To overcome this issue, we propose a conditional prior specification for the split variables that favors model parsimony.

1 Population MCMC with efficient tree proposals

Population MCMC methods (Geyer, 1991) allow both local and global transitions by simulating a number of auxiliary MCMC chains targeting progressively tempered posterior densities, and swapping their states with probability $\xi$. The Markov chain targeting the untempered posterior can thus explore greatly separated regions of the parameters space efficiently relative to a single chain. In Algorithm 1 we incorporate the proposal of Pratola (2016) in a parallel tempering sampler targeting the posterior distribution $[\tau, \{(v_i, c_i)\}, \sigma^2, \mu \mid y]$. Symbols with subscripts in parentheses correspond to a single chain. We use an expanded notation for the likelihood to make clear the dependence on sampled parameters at each step. The user defines the vector of temperatures $\gamma \in (0, 1)^C$.

*Main article DOI: 10.1214/16-BA999.

1Department of Statistics, The Ohio State University, Columbus, OH, USA, oksana@stat.osu.edu
Algorithm 1 Parallel tempering tree sampling algorithm using \( C \) Markov chains.

For \( c = 1, \ldots, C \) construct \( T_c = (\tau_c, \{(v_i, c_i)\}_{c}) \) using the algorithm of Chipman et al. (2010), then sample \( \sigma_c^2, \mu_c \) from the conditional prior;

for \( m = 1 : M_c \) do
  if \( \xi_m > U[0,1] \) then
    Propose a swap between the index pair \((i, j)\) drawn from a symmetric proposal distribution \( g(i, j) \), \( 1 \leq i, j \leq C, i \neq j \) and compute the ratio,
    \[
    \rho = \frac{L(y | \tau_{ij}, \{(v_i, c_i)\}_{ij}, \sigma_{ij}^2, \mu_{ij})}{L(y | \tau_j, \{(v_i, c_i)\}_j, \sigma_j^2, \mu_j)} \frac{L(y | \tau_{ij}, \{(v_i, c_i)\}_{ij}, \sigma_{ij}^2, \mu_{ij})}{L(y | \tau_j, \{(v_i, c_i)\}_j, \sigma_j^2, \mu_j)} \gamma_j;
    \]

  if \( \min(1, \rho) > U[0,1] \) then
    Swap \((\tau_{ij}, \{(v_i, c_i)\}_{ij}, \sigma_{ij}^2, \mu_{ij}) \leftrightarrow (\tau_j, \{(v_i, c_i)\}_j, \sigma_j^2, \mu_j)\);
  end if
end if

for \( c = 1 : C \) do
  if \( \xi_r > U[0,1] \) then
    Construct \( T'_c = (\tau'_c, \{(v'_i, c'_i)\}_{c}) \) by performing a birth/death or rotation on \( T_c = (\tau_c, \{(v_i, c_i)\}_{c}) \) and compute the tempered ratio:
    \[
    \rho = \frac{\pi(T_c)p_c(T_c)p^1_c p^2_c L(y | \tau_c, \{(v_i, c_i)\}_c, \sigma_c^2, \mu_c) \gamma_c}{\pi(T'_c)p_c(T'_c)p^1_c p^2_c L(y | \tau'_c, \{(v'_i, c'_i)\}_c, \sigma'_c, \mu'_c) \gamma_c};
    \]

  if \( \min(1, \rho) > U[0,1] \) then
    Update \( T_c \leftarrow T'_c \);
  end if
end if

if \( \xi_p > U[0,1] \) then
  For \( i = 1, \ldots, |T_c| \), with probability proportional to equation (7), propose \((v'_i, c'_i) \) by performing a perturb or perturb within change-of-variable proposal \( q \) in equation (6), and compute the tempered ratio:
    \[
    \rho = \frac{\pi((v_i, c_i) \mid c_i) q(v'_i, c'_i \mid v_i, c_i) L(y | v_i, c_i) \gamma_c}{\pi((v'_i, c'_i) \mid c_i) q(v_i, c_i \mid v'_i, c'_i) L(y | v'_i, c'_i) \gamma_c};
    \]

  if \( \min(1, \rho) > U[0,1] \) then
    Update \((v_i, c_i) \mid c_i \leftarrow (v'_i, c'_i) \mid c_i\);
  end if
end if

For \( j = 1, \ldots, M_c \), draw \( \mu_{j(c)} \) from \( \mu_j \mid y, \tau(c), \{(v_i, c_i)\}_{c}, \sigma^2(c) \)

Draw \( \sigma^2(c) \) from \( \sigma^2 \mid y, \tau(c), \{(v_i, c_i)\}_{c}, \mu(c) \)

end for

Save the state, \((\tau(c), \{(v_i, c_i)\}_{c}, \sigma^2(c), \mu(c))\), of the final chain.
end for

with the last element equal to one, and the probabilities \( \xi_r \) and \( \xi_p \) of rotating a single tree or perturbing a single variable, respectively. Note that the sampling over all non-swap candidate chains can be performed in parallel at every iteration.
2 Prior for model selection

In the context of estimation, the prior of Chipman et al. (2010) acts against over-fitting the data by penalizing the depth of internal nodes while putting uniform weights on the possible indices, \(\{1, \ldots, d\}\), of the split variables \(v\). For the problem of model selection we instead suggest a prior that encourages parsimony, penalizing the number of distinct variables along which splits are made. This can be accomplished by introducing prior dependence among the split variables \(v\). We may define the prior on \(v_i\) conditionally such that its distribution should put most of its mass on the unique values of all the less deep nodes \(v_1, \ldots, v_{i-1}\) with the remaining prior mass uniformly distributed among all variables on which splits have not yet been made. As with the prior of Chipman et al. (2010), draws from the proposed model selection prior can be constructed sequentially and have a convenient closed form. As suggested in the first motivating example, this prior choice may result in more sharply defined posterior modes in cases when variables are confounded. However, the very efficient proposal strategy of Pratola (2016), combined with particle MCMC may nevertheless be able to identify these modes.

References


