Technical Report #136

Bayesian estimation of free-knot splines using reversible jumps

Mary Lindstrom, PhD
Bayesian estimation of free-knot splines using reversible jumps

Mary J. Lindstrom

September, 1999

University of Wisconsin–Madison
Department of Biostatistics & Medical Informatics
Technical Report #136

Key words: conditionally linear parameters, data based prior, reversible jump Markov chain
Monte Carlo

1This research was supported in part by National Institutes of Health grants Nos. CA75097 and DC00820
Abstract

A fully Bayesian approach to estimating cubic free-knot splines is described. A new transition kernel for a reversible jump Markov chain Monte Carlo sampler is developed including a general method for constructing proposals for conditionally linear parameters. A general prior for the knots is proposed which allows a varying amount of prior probability on knot vectors with nearly identical knots. A data based prior is used for the conditionally linear coefficients which avoids the tendency to assign all posterior probability on the smallest model when the range of the coefficients is large compared to the variability in the data. Dimension changing moves include moves to increase/decrease the knot vector by an arbitrary number of knots which improves mixing; particularly when the posterior for the number of knots is multi-modal. We apply the method to two data sets.
1 Introduction

A regression spline on the interval \([\gamma_0, \gamma_{k+1}]\) with ordered knots \(\gamma = (\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_k, \gamma_{k+1})\), is a polynomial of order \(m\) (degree \(m - 1\)) between each pair of adjacent knots. If the knots are distinct, then the polynomial pieces are joined so that the first \(m - 2\) derivatives of the spline are continuous. Most often, the order is taken to be 4 since the resulting cubic spline is smooth to the eye when the knots are distinct. We choose to restrict the model to the space of cubic natural splines which improves the behavior of the estimate at the endpoints of the design interval by placing restrictions on the derivatives. Specifically we fix the knots \(\gamma_0\) and \(\gamma_{k+1}\) at the left and rightmost design point and set the second derivative to zero at \(\gamma_0\) and \(\gamma_{k+1}\). A cubic natural spline regression model can be written in the form

\[
y = \sum_{j=1}^{k+2} c_j B_j(x, \gamma) + \epsilon
\]

where \(y = [y_1, \ldots, y_n]\) is the observed data, \(x = [x_1, \ldots, x_n]\) is the vector of design points, \(c = [c_1, \ldots, c_{k+2}]\) is a vector of coefficients, \(B_j, j = 1, \ldots, k + 2\) is a set of functions that form a basis for the space of cubic natural splines with knots \(\gamma\), and \(\epsilon = [\epsilon_1, \ldots, \epsilon_n]\) is a noise term with distribution centered at zero and variance parameters \(\alpha\).

The spline coefficients \(c\) are conditionally linear parameters (given the knots \(\gamma\)) and, if \(\gamma\) is known, can be easily estimated via linear least squares. If the number and location of the knots are unknown (a more realistic assumption) then Equation 1 is a non-parametric regression model called a free-knot spline. Free-knot splines are attractive because the fitted curve is a reasonably simple, low dimensional, analytic function. Also, smoothing is local and when the knots are allowed to range over the whole design interval the fitted curve can include non-smooth features. However, there are a number of practical difficulties in constructing parameter estimates and assessing the variability of the fitted curve. First, an estimator for \(\gamma\) must be defined for a given \(k\). (Note that our definition of \(\gamma\) includes fixed
end knots $\gamma_0$ and $\gamma_{k+1}$ which, of course, do not enter into the estimation problem.) Second, $k$ must be estimated which usually requires computing an estimate of $\gamma$ for each reasonable value of $k$. Lastly the variability of the fitted curve must be assessed including uncertainty in $k$.

The most commonly discussed estimation method for $\gamma$ given $k$ is nonlinear least squares (de Boor and Rice, 1968; Jupp, 1978; Lindstrom, 1999). However, this estimator often contains unnecessary non-smooth features. Lindstrom (1999) proposes a penalized estimator that improves the computational properties of free-knot splines by penalizing knot vectors with knots that are close together. This estimator eliminates unnecessary “corners” in the fitted curve, and in simulation studies, shows no increase in the loss.

We can address the issues of estimating $k$ and accounting for uncertainty in $k$ by casting the problem in a Bayesian framework where we are interested in functions of the posterior distribution

$$P(k, \theta | y) = P(k|y)P(\theta | y, k)$$

where $\theta = [\gamma \ c \ \alpha]$. Knot vectors with nearly replicate knots can be down weighted by using a Dirichlet prior on the inter-knot spacings (See Section 4.3 for details).

There is no general analytical expression for the posterior distribution but methods exist to construct a sample from it. Our choice of methods is limited by the fact that the dimension of the model is unknown. The approach we will use is reversible jump Markov chain Monte Carlo (RJMCMC) described in Green (1995) and Richardson and Green (1997).

## 2 Reversible Jump MCMC

Reversible jump MCMC is a generalization of the Metropolis-Hastings (M-H) sampler (Tierney, 1994) that allows moves that increase or decrease the dimension of the parameter vector.
The transition kernel for RJMCMC is usually constructed using proposals of the form

$$\begin{align*}
[\theta' \ u_2] &= f([\theta \ u_1])
\end{align*}$$

where $\theta$ is the current state, $\theta'$ is the proposed state, $u_1$ and $u_2$ are random variables (possibly dependent on the current state) and $f$ is a deterministic invertible function. The invertibility of $f$ implies “dimension matching”, i.e.

$$\text{length}(\theta) + \text{length}(u_1) = \text{length}(\theta') + \text{length}(u_2)$$

and it is typical to define the reverse move as $[\theta \ u_1] = f^{-1}([\theta' \ u_2])$. Also, for moves that change dimension, either $u_1$ or $u_2$ will often have length 0.

The acceptance probability for RJMCMC has the form $\text{min}(1, A)$ where

$$A = \frac{P(y|\theta', k) \times P(k')P(\theta'|k') \times q(\theta|\theta') \times \abs{\frac{\partial f([\theta \ u_1])}{\partial [\theta \ u_1]}}}{P(y|\theta, k) \times P(k)P(\theta|k) \times q(\theta|\theta) \times \abs{\partial f([\theta \ u_1])}}$$

where

$$\begin{align*}
\frac{q(\theta'|\theta)}{q(\theta|\theta')} &= \frac{q(M|\theta)q(i|\theta, M)q(u_2|\theta, M, i)}{q(M'|\theta')q(i'|\theta', M')q(u_1|\theta', M', i')} \\
\end{align*}$$

where $M$ indicates the move type and $i$ indicates the particular items to modify ($i$ may be multivariate). Often this will simplify to

$$\begin{align*}
\frac{q(\theta'|\theta)}{q(\theta|\theta')} &= \frac{q(M|k)q(i|M, k)q(u_2|\theta, M, i)}{q(M'|k')q(i'|M', k')q(u_1|\theta', M', i')}
\end{align*}$$

In the M-H sampler the determinant of the Jacobian is 1, $k = k'$, $u_1 = 0'$, $u_2 = 0$, $f$
simply shuffles the order of the parameters and $A$ simplifies to the familiar form

$$A_{\text{M-H}} = \frac{P(y|\theta')}{P(y|\theta)} \frac{P(\theta')}{P(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$$

### 3 MCMC for free-knot splines

Regardless of the type of MCMC sampler used, moves that change the knot vector in a free-knot spline must also update the coefficient values to produce a proposed state with a reasonable chance of being accepted. One way to avoid this issue is to use a profile likelihood $P(y|\gamma, \hat{c}(\gamma, y), \alpha, k)$ where the coefficients are set equal to their least squares estimates given a particular knot vector. This is the approach taken by Denison et al. (1998) for a piecewise polynomial model. While the posterior mode will likely not be effected by the use of the profile likelihood, the posterior will not reflect the variability due to the estimation of the coefficients which can be considerable if the signal to noise ratio is not large.

Since we wish to use the full likelihood we must update the coefficients whenever the knot vector is changed. This process is dependent on the set of basis functions used to construct the model. Typically the truncated power basis is used when the coefficients must be updated (Denison et al., 1998; Green, 1995). Using this basis, the model has the form

$$\sum_{j=0}^{m-1} b_j x^j + \sum_{j=1}^{k} b_{j+m}(x - \gamma_j)^m - 1$$

where $(t)_+ = t$ if $t > 0$ and 0 otherwise. This basis makes it easy to construct a reasonable proposal for the coefficient vector when the knot vector is modified. However, if we allow the knots to vary over the whole design interval the truncated power basis is computationally quite unstable. As two knots move together the least squares estimates of the coefficients are very poorly determined. While we penalize knots that are close together we do not wish to exclude them entirely since for some examples non-smooth features are appropriate.
The b-splines basis (de Boor, 1978) is the most computationally convenient and stable basis for standard splines. It is well defined even for knots vectors with replicate knots and there exist at least two related bases for the natural splines with similar properties. However, there is no obvious way to construct a new coefficient vector when the knot vector is modified. There do exist methods for converting between the truncated power and b-spline bases (de Boor, 1978), but the methods are quite complex and give no insight into constructing new coefficient vectors for the b-splines.

We use the restricted b-spline basis for the natural splines (see Appendix A) for its computational stability and solve the coefficient updating problem by first updating the “coefficient error” \( e = c - \hat{c}(\gamma, y) \) where \( \hat{c}(\gamma, y) \) is the least squares estimate of the coefficients given the data and the current knots. That is, we propose \( c' = \hat{c}(\gamma', y) + e' \) where \( e' \) is a modified version of \( e \). Proposing \( e' \) from \( e \) when, for example, a knot is added need not be more complicated than inserting a new element generated from a mean zero distribution. More formally, we create proposal \( \gamma' \) and \( c' \) as

\[
[\gamma' \ v_2 \ c' \ w_2] = f_i([\gamma \ v_1 \ c \ w_1])
\]

where \( f_i \) is an invertible function chosen so as to create a proposal that has reasonable likelihood of being accepted and where the transformation \( f_i \) is defined using the component transformations:

\[
\begin{align*}
  e &= c - \hat{c}(\gamma, y) \\
  [e' \ w_2] &= g_{e,i}([e \ w_1]) \\
  [\gamma' \ v_2] &= g_{\gamma,i}([\gamma \ v_1]) \\
  c' &= \hat{c}(\gamma', y) + e'
\end{align*}
\]

(2)

Note that \( e \) and \( e' \) are intermediate variables used to make the exposition clearer and not additional parameters in the model.

This approach will work for any model with conditionally linear parameters and has the
added advantage of not increasing the complexity of the Jacobian for moves that change dimension. The Jacobian has the form

$$
\begin{vmatrix}
\frac{\partial [\gamma', c']}{\partial [\gamma v c w]}
\end{vmatrix} = \begin{vmatrix}
\frac{\partial [\gamma v_2]}{\partial [\gamma v_1]} & \frac{\partial [\gamma' v_2]}{\partial [c w_1]}
\frac{\partial [c' w_2]}{\partial [\gamma v_1]} & \frac{\partial [c' w_2]}{\partial [c w_1]}
\end{vmatrix} = \begin{vmatrix}
\partial g_{x,i}((\gamma v_1)) & \partial g_{x,i}((c - \hat{c}(\gamma, y) w_1))
\partial [\gamma v_1] & \partial [c w_1]
\end{vmatrix}
$$

(3)

since \((\partial [\gamma' v_2])/(\partial [c w_1]) = 0\). For the reverse move we construct the knot and coefficient vectors through the inverse transformation \([\gamma' v_1, c' w_1] = f_i^{-1}(\gamma v_2 c w_2)\). We apply this approach in the following section for moves that change the location of a knot and moves that increase or decrease the number of knots.

4 Implementation

4.1 Moves that do not change dimension

Moves that do not change the dimension of the parameter vector are implemented exactly as in a M-H sampler. Table 1 describes the proposal mechanisms and the values of the proposal ratios for four move types that do not change the dimension of the parameter vector. The first move type is a local change to a knot value. The change is local in that the knot \(\gamma_i\) remains in the interval \([\gamma_{i-1}, \gamma_{i+1}]\). We use a triangular distribution for the new knot centered at the knot it is replacing. A distribution with a scale parameter could be used instead to gain control over the acceptance probabilities for the move. The second move type is a global change to a knot where the proposal can range over the entire design interval \([\gamma_0, \gamma_k + 1]\) (thus possibly renumbering the existing knots). Local knot moves will have a higher acceptance rate than global but global moves improve mixing. For both knot move types, the coefficient vector is updated by holding \(e\) constant and setting \(c' = \hat{c}(\gamma', y) + e\).

The third move type is a change to a coefficient (with no change to the knot vector). These
Table 1: Proposal probabilities for moves that do not change the number of knots

<table>
<thead>
<tr>
<th>Probability of choosing move type</th>
<th>knot local</th>
<th>knot global</th>
<th>coefficient</th>
<th>sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_f p_g k$</td>
<td>$p_f (1 - p_g) k$</td>
<td>$p_f (k + 2)$</td>
<td>$p_f$</td>
<td>$2k + 3$</td>
</tr>
</tbody>
</table>

Proposal mechanism given move type

| $i \sim U_d(1, \ldots, k)$ | $i \sim U_d(1, \ldots, k)$ | $i \sim U_d(1, \ldots, k + 2)$ | $v \sim U(-\delta, \delta)$ |
| $\gamma_i' \sim U(\gamma_0, \gamma_{i+1})$ | $\gamma_i' \sim U(\gamma_0, \gamma_{i+1})$ | $\gamma_i' \sim U(\gamma_0, \gamma_{i+1})$ | $\gamma_i' \sim U(\gamma_0, \gamma_{i+1})$ |
| $e' = c - \gamma_i' + \varepsilon$ | $e' = c - \gamma_i' + \varepsilon$ | $e' = c - \gamma_i' + \varepsilon$ | $e' = c - \gamma_i' + \varepsilon$ |

Proposal ratio

| $d_T(\gamma_i, \gamma_i', \gamma_i-1, \gamma_i+1)$ | $d_T(\gamma_i, \gamma_i', \gamma_i-1, \gamma_i+1)$ | $1$ | $1$ | $\sigma'/\sigma$ |

$p_f$ is the probability of choosing a move that keeps the dimension of the parameter vector fixed, $p_g$ is the probability of choosing to move a knot globally rather than locally, $U(a, b)$ is a uniform distribution on the interval $[a, b]$, $U_d(a)$ is a discrete distribution with probability $1/\text{length}(a)$ on the entries of $a$, $T(H, a, b)$ denotes a triangular distribution with peak $H$, range $[a, b]$ and density $d_T(x, H, a, b)$.

Moves are proposed as the current coefficient plus a normal perturbation. In our examples we assume $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and the fourth move type proposes updates to $\sigma$ uniformly in the log scale. See Section 5 for details on setting the tuning constants for the proposal distributions.

4.2 Moves that change dimension

Add/delete move pair

The simplest increase/decrease move pair is to add/delete a single knot. For the add move we first choose an interval $i$ in which to place the new knot $i \sim v_d(1, \ldots, k + 1)$. From Equation 2 we can define the move by defining updates to $\gamma$ and $\varepsilon$. We use the very simple
form:

\[
\begin{align*}
\gamma' &= g_{\gamma,d}(\gamma v) = [\gamma_1, \ldots, \gamma_{i-1}, v, \gamma_i, \ldots, \gamma_k] \\
\epsilon' &= g_{\epsilon,d}(\epsilon w) = [\epsilon_1, \ldots, \epsilon_i, w, \epsilon_{i+1}, \ldots, \epsilon_{k+2}]
\end{align*}
\]

where \( v \sim U(\gamma_{i-1}, \gamma_i) \) and \( w \sim \mathcal{N}(0, \sigma^2_w) \). For the corresponding delete move we let \( i \sim U_d(1, \ldots, k) \) define which knot to delete. The move is then determined by inverting the add move. That is

\[
\begin{align*}
\gamma' &= [\gamma_1, \ldots, \gamma_{i-1}, \gamma_{i+1}, \ldots, \gamma_k] \\
v &= \gamma_i \\
\epsilon' &= [\epsilon_1, \ldots, \epsilon_i, \epsilon_{i+2}, \ldots, \epsilon_{k+2}] \\
w &= \epsilon_{i+1}
\end{align*}
\]

For the add move the proposal ratio is

\[
\frac{(k + 1)\eta_D(k + 1)}{(k + 1)\eta_A(k)q_v(v)q_w(w)} = \frac{\eta_D(k + 1)(\gamma_i - \gamma_{i-1})}{\eta_A(k)d_N(w, 0, \sigma^2_w)}
\]

where \( \eta_A(k) \) and \( \eta_D(k) \) are the probabilities of choosing add and delete moves respectively when the number of free knots is equal to \( k \) and \( d_N(x, \mu, \sigma^2) \) is the normal density with mean \( \mu \) and variance \( \sigma^2 \) evaluated at \( x \). From Equation 3 the determinant of the Jacobian reduces to

\[
\begin{vmatrix}
\frac{\partial[\gamma', \gamma_{i+1}]}{\partial[\gamma_i, v]} & \frac{\partial[\epsilon', \epsilon_{i+1}, \epsilon_{i+2}]}{\partial[\epsilon_i, c_{i+1}, w]}
\end{vmatrix} = \begin{vmatrix}
\frac{\partial[v, \gamma_i]}{\partial[\gamma_i, v]} & \frac{\partial[\epsilon_i, w, \epsilon_{i+1}]}{\partial[c_i, c_{i+1}, w]}
\end{vmatrix} = 1
\]

For the delete move the proposal ratio is \( \eta_A(k - 1)d_N(w, 0, \sigma^2_w)/[\eta_D(k)(\gamma_{i+1} - \gamma_{i-1})] \) and the determinant of the Jacobian is also 1.

Note that the add move that we have defined makes insertion of a knot in each interval equally likely regardless of the length of the interval. An alternative add move could be constructed choosing a new knot with uniform probability over \([\gamma_0, \gamma_{k+1}]\).
Split/Combine move pair

A second increase/decrease move that we will use is splitting a knot into two and the reverse: combining two knots into one. For the split move we first choose a knot \( i \) to split \( i \sim U_d(1, \ldots, k) \), and define two new knots to take its place as

\[
\gamma' = g_{\gamma,i}(\gamma, v) = [\gamma_1, \ldots, \gamma_{i-1}, \gamma_i - v(\gamma_i - \gamma_{i-1}), \gamma_i + v(\gamma_{i+1} - \gamma_i), \gamma_{i+1}, \ldots, \gamma_k]
\]

where the density for \( v \) is \( 2 - 2v \) for \( 0 < v < 1 \), and 0 otherwise. This choice of density for \( v \) makes splits that stay closer to the knot being replaced more likely. Other reasonable choices for \( v \) include a uniform \((0,1)\) random variable. The proposed coefficient error vector is

\[
e' = g_{e,i}(\gamma, w) = [e_1, \ldots, e_i, e_{i+1} - w, e_{i+1} + w, e_{i+2}, \ldots, e_{k+2}]
\]

where \( w \sim \mathcal{N}(0, \sigma_w^2) \).

For the corresponding combine move, assume there are currently \( k \) free knots and let \( i \sim U_d(1, \ldots, k - 1) \). The proposal that inverts the split move combines the \( i \)th and \( i + 1 \)st knots as

\[
\gamma' = [\gamma_1, \ldots, \gamma_{i-1}, \gamma_i - (\gamma_i - \gamma_{i-1})/(1 - v), \gamma_{i+2}, \ldots, \gamma_k]
\]

\[
v = (\gamma_{i+1} - \gamma_i)/(\gamma_{i+2} - \gamma_{i-1})
\]

\[
e' = [e_1, \ldots, e_i, (e_{i+1} + e_{i+2})/2, e_{i+3}, \ldots, e_{k+2}]
\]

\[
w = (e_{i+1} - e_{i+2})/2
\]

Note that \( \gamma_i' \) will lie between \( \gamma_{i-1}' \) and \( \gamma_{i+1}' \) and \( v \) between 0 and 1. The proposal ratio for the split move is

\[
\frac{k \eta_C(k + 1)}{k \eta_S(k) q_v(v) q_w(w)} = \frac{\eta_C(k + 1)}{\eta_S(k) (2 - 2v) d_N(w, 0, \sigma^2_w)}
\]

where \( \eta_S(k) \) and \( \eta_C(k) \) are the probabilities of choosing a split and combine move respectively given that the current knot vector has \( k \) free knots. The determinant of the Jacobian
reduces to
\[
\left| \frac{\partial [\gamma_{i-1}', \gamma_i', \gamma_{i+1}', \gamma_{i+2}']}{\partial [\gamma_{i-1}, \gamma_i, \gamma_{i+1}, v]} \right| = (1 - v)(\gamma_{i+1} - \gamma_{i-1})
\]

The proposal ratio for the combine move is \(\eta_S(k-1)(2 - 2u) d_N(w, 0, \sigma_w^2) / \eta_C(k)\) and the determinant of the Jacobian is the inverse of that for the split move with a change of variables: \([(1 - v)(\gamma_{i+2} - \gamma_{i-1})]^{-1}.

**Add 2/delete 2 move pair**

Because of the multi-modal nature of the likelihood surface for free-knot splines (Jupp, 1978; Lindstrom, 1999) a move pair that increases/decreases the number of knots by two will likely increase mixing. It is simplest to implement this as an add/delete pair. For the “add two knots” move we choose two intervals \(i\) and \(j\) in which to place the new knots where \(i\) and \(j\) are both distributed \(U_d(1, \ldots, k + 1)\). We simulate \(v_1 \sim U(\gamma_{i-1}, \gamma_i)\), \(v_2 \sim U(\gamma_{j-1}, \gamma_j)\) and \(w_1\) and \(w_2\) both distributed \(N(0, \sigma_w^2)\). We construct \(\gamma'\) by inserting the new knots \(v_1\) and \(v_2\) into intervals \(i\) and \(j\) and construct \(e'\) by inserting \(w_1\) between \(e_i\) and \(e_{i+1}\) and inserting \(w_2\) between \(e_j\) and \(e_{j+1}\). If \(i = j\) the order of \(w_1\) and \(w_2\) is determined by the order of \(v_1\) and \(v_2\). For the corresponding delete move we let \(i \sim U_d(1, \ldots, k)\) and \(j \sim U_d(1, \ldots, i - 1, i + 1, \ldots, k)\) define the knots to delete. The move is then determined by inverting the add move. That is, \(\gamma'\) is \(\gamma\) with the \(i\)th and \(j\)th knots removed and \(e'\) is \(e\) with the \(e_{i+1}\) and \(e_{j+1}\) removed, \(v = [\gamma_i \gamma_j]\), and \(w = [e_{i+1} e_{j+1}]\).

For the “add two knots” move the proposal ratio is
\[
\frac{(k + 2)^{-1}(k + 1)^{-1}\eta_D(k + 2)(\gamma_i - \gamma_{i-1})(\gamma_j - \gamma_{j-1})}{(k + 1)^{-2}\eta_A(k)d_N(w_1, 0, \sigma_w^2)d_N(w_2, 0, \sigma_w^2)}
\]

where \(\eta_A(k)\) and \(\eta_D(k)\) are the probabilities of choosing the “add two knots” and “delete two knots” moves respectively when the number of free knots is equal to \(k\). For the corre-
sponding delete move the proposal ratio is

\[
\frac{(k - 1)^{-2} \eta_{A2}(k - 2) d_{\mathcal{N}}(w_1, 0, \sigma_w^2) d_{\mathcal{N}}(w_2, 0, \sigma_w^2)}{k^{-1}(k - 1)^{-1} \eta_{D2}(k) g(i, j, \gamma)}
\]

where \(g(i, j, \gamma) = (\gamma_{i+1} - \gamma_i - 1)(\gamma_{j+1} - \gamma_j - 1)\) if \(i, j\) are not contiguous and \((\gamma_{i^*+2} - \gamma_{i^*-1})^2\) if \(i\) and \(j\) are contiguous where \(i^* = \min(i, j)\). The determinant of the Jacobian for both moves is 1.

This approach to constructing add/delete moves can be easily extended to move pairs that increase/decrease by 3 or more. We use moves that add/delete up to 3 knots for the examples in Section 5.

### 4.3 Priors

We use a Poisson prior with parameter \(\lambda = 1\) for the number of knots. This gives decreasing prior probability as \(k\) increases which keeps the model size moderate. It is possible to keep the model size moderate by increasing the dispersion in the priors on the parameters that change dimension but we choose instead to use the prior for \(k\) for this purpose.

We use uninformative priors for the remainder of the parameters. As discussed in Section 1 we use a Dirichlet prior for the knot vector. We assume \(h(\gamma) \sim D(\alpha 1)\) where \(h(\gamma) = (\gamma_1 - \gamma_0, \gamma_2 - \gamma_1, \ldots, \gamma_{k+1} - \gamma_k)/(\gamma_{k+1} - \gamma_0)\) and \(1\) is a unit vector of length \(k + 1\). When \(\alpha = 2\) this prior is equivalent to the order statistic prior use by Green (1995) for the coal mining disasters example. We used \(\alpha = 1.1\) which corresponds to less penalty on knots that are close together. The result for the examples in Section 5 are not sensitive to the value of \(\alpha\).

Because the range of the coefficients may be large compared to the variability in the data, specifying an uninformative prior for the coefficients must be done with care to avoid placing all the posterior probability on the smallest models. We take advantage of the
<table>
<thead>
<tr>
<th>Proposal Parameters</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_f$ proportion of proposed moves that do not change $k$</td>
<td>.5 .5</td>
</tr>
<tr>
<td>$p_g$ proportion of knot moves ($k$ fixed) that are global</td>
<td>.4 .4</td>
</tr>
<tr>
<td>$\sigma_1$ std. dev. of proposed coefficient (chosen to make acceptance proportion $\approx 0.5$)</td>
<td>.25 .1</td>
</tr>
<tr>
<td>$\delta$ variability of proposed $\sigma$ (chosen to make acceptance proportion $\approx 0.5$)</td>
<td>.3 .1</td>
</tr>
<tr>
<td>$\sigma_w$ std. dev. of proposed coefficient errors for moves which change $k$ (chosen to make acceptance proportion $\approx 0.5$)</td>
<td>$\sigma_1$ $\sigma_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$ Poisson parameter for $k$</td>
<td>1 1</td>
</tr>
<tr>
<td>$\alpha$ Dirichlet parameter for $\gamma$</td>
<td>1.1 1.1</td>
</tr>
<tr>
<td>$\sigma_c$ standard deviation of $c$</td>
<td>$\sigma_1$ $\sigma_1$</td>
</tr>
<tr>
<td>$\tau$ std. dev. parameter in half-normal distribution of $\sigma$ (set equal to 5 times the standard deviation of the data)</td>
<td>1.25 1.5</td>
</tr>
</tbody>
</table>

Table 2: Parameter values

The conditional linearity of the coefficients given the knots and specify a data dependent prior $c \sim \mathcal{N}(\hat{c}(\gamma, y), \sigma_c)$. We use $\sigma_c = \sigma_1$ (see Table 2) which results in an observed standard deviation of the elements of $e = c - \hat{c}(\gamma, y)$ of approximately $\sigma_c/2$. Thus the variability in the coefficients is not unduly constrained. We specify a half normal prior distribution for $\sigma$ with standard deviation parameter $\tau$ which we set large relative to the standard deviation of the data. A disperse prior for $\sigma$ will not concentrate the posterior distribution of $k$ on the smallest model since the dimension of $\sigma$ is fixed.

5 Examples

We used the RJMCMC methodology described above to obtain a sample from the posterior distributions for each of two examples. For both examples we allow add/delete moves that increase/decrease the knot vector by up to 3 knots. The probability of choosing each of the
Figure 1: The voltage drop data with fitted curves. The curved dotted and dashed lines are the fitted curves corresponding to the state with largest posterior probability for $k = 4$ and 5. The 4 knot solution is the estimated posterior mode. The knot locations are shown below the data as triangles on the straight dotted and dashed lines. The light grey curves correspond to 50 randomly selected curves from the MCMC sample.

8 moves that change dimension is set to $(1 - p_f)/8$ unless one or more of the move types are impossible for the current number of knots in which case those move types have probability zero and the remaining dimension switching move probabilities are scaled accordingly. We ran both chains for 100,000 updates saving every 10th update after the first 10,000.

The first example is data on the voltage drops in the battery of a guided missile motor given in Eubank (1988) and shown in Figure 1. Table 2 gives details on the parameter values used for the transition kernel and the priors for this and the following example. Note that we determined $\sigma_1$ using preliminary runs of the sampler and then used this preliminary estimate to define $\sigma_c$ and $\sigma_w$. The sequence of saved states for $k$ for this example is displayed in the top panel of Figure 2 and the posterior for $k$ (a simple frequency plot) is shown in the top panel of Figure 3.
Figure 2: The series of saved states for the parameter $k$, the number of interior knots.

Figure 3: Estimated posterior distributions of $k$, the number of interior knots, for the voltage drop data and the simulated example.
Figure 4: Plot of 500 simulated data points, the true curve (solid line), and 2 and 5 knot fitted curves (dotted and dashed lines). The light grey curves correspond to 50 randomly selected curves from the MCMC sample. See the caption for Figure 1 for further details.

The 4 and 5 knot fits to the voltage drop data are shown in Figure 1. The 4 knot curve corresponds to the posterior mode. That is, it corresponds to the state with the highest posterior probability for the value of $k$ with highest posterior probability. We use the highest posterior modes rather than means for our estimates because of the multi-modal nature of the posterior. The light grey curves corresponding to 50 random draws from the MCMC sample and give a sense of the variability around the mode.

The simulated data example is plotted with the true curve and 2 and 5 knot fitted curves in Figure 4. The estimated posterior for $k$ for this example (bottom panel, Figure 3) is particularly interesting because it is bimodal with the majority of the posterior probability centered at $k = 5$ and a second smaller mode at $k = 2$. The 50 randomly drawn curves in Figure 4 reflect this bimodality. Move pairs that increase/decrease the number of knots by more than 1 were particularly important for this example even though their acceptance
percentages were low (Table 3). Without these move types, the “mode jumping” from k=5 to k=2 shown in the bottom panel of Figure 2 is much less frequent.

Diagnostics for RJMCMC are not well developed. The percentage of dimension changing moves accepted is one indication of the mixing of the chain, however, there is no rule as to what this percentage should be. Richardson and Green (1997) report values from 4 to 18%. Our acceptance rates range from 1 to 28% (Table 3).

Traditional MCMC diagnostics can be used for parameters that do not change meaning as the dimension changes and for predicted values. We used time series plots, sample autocorrelation functions, and CUSUM plots to assess the stationarity of the chain for \( \sigma \) and for ten equally space predicted values. All of these diagnostics indicated acceptable stationarity and good mixing. In addition, chains started at various over-dispersed states converged to the same stationary distribution. The standard errors of the ten predicted values (Kosorok, 1999) were less than 0.003 for the voltage drop data and less than 0.005 for the simulated example indicating good effective sample size.

### Discussion

We have describe a fully Bayesian approach to estimating a free knot spline using RJMCMC and applied it to two examples. One of the most important advantages of the Bayesian approach is that the resulting sample from the posterior provides excellent information about the uncertainty in the fitted curve which includes the uncertainty in the number of

<table>
<thead>
<tr>
<th>Example</th>
<th>( \gamma ) local</th>
<th>( \gamma ) global</th>
<th>( c )</th>
<th>( \sigma )</th>
<th>Split</th>
<th>Combine</th>
<th>Add</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Delete</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage drop</td>
<td>53</td>
<td>34</td>
<td>47</td>
<td>55</td>
<td>12</td>
<td>28</td>
<td>17</td>
<td>12</td>
<td>3</td>
<td>21</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated</td>
<td>55</td>
<td>21</td>
<td>38</td>
<td>48</td>
<td>10</td>
<td>9</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Move acceptance percentages
knots. The RJMCMC sampler is particularly well suited to free knot splines because knot vectors which produce good fits often contain knot subsets which also produce good fits. While the method requires the specification of a number of parameters in the proposal and prior distributions, we have provided a simple recipe for their choice (Table 2).

New features of our sampler include a transition kernel that takes advantage of the conditional linearity of the spline coefficients to obtain good acceptance percentages for moves that change dimension. Also a data based prior for the coefficients is constructed that avoids the tendency to place all the posterior probability on the smallest model. Moves that add/delete up to three knots are used to improve mixing and the extension to larger jumps in the number of knots is straightforward. We also use a prior for the knot locations that allows us to vary the amount of penalty on knot vectors with knots that are close together.

Appendix A

There are a number of choices for the basis functions for a natural spline. We use a normalized restricted b-spline basis derived from the basis implemented in the S-plus function nsbasis(). The restricted b-spline basis for cubic natural splines with knots $\gamma$ is constructed by first calculating the $n$ by $k + 4$ matrix of b-splines $B$ evaluated at the $n$ design points. The space of cubic splines is spanned by the columns of $B$ and any cubic spline can be represented as $Bc$. We can restrict the space to the cubic natural splines by requiring $Dc = 0$ where $D$ is the 2 by $k + 4$ matrix of 2nd derivatives of $Bc$ with respect to $x$ evaluated at $x = \gamma_0$ and $x = \gamma_{k+1}$. If we let $QR$ be the QR-decomposition (Kennedy and Gentle, 1980) of $D^T$ then we have

$$Bc = BQQ^Tc = \begin{bmatrix} BQ_1 & BQ_2 \end{bmatrix} \begin{bmatrix} 0 \\ Q_2^Tc \end{bmatrix} = \begin{bmatrix} BQ_2 \end{bmatrix}Q_2^Tc$$
where $Q_1$ and $Q_2$ are the first 2 and last $n-2$ columns of $Q$ respectively. This implies that we can represent any spline of the form $Bc$ for which $Dc = 0$ as the product of an $n$ by $k+2$ matrix $BQ_2$ and a vector $Q_2^Tc$ of length $k+2$. Thus $BQ_2$ is a basis for the cubic natural splines with knots $\gamma$.

We further modify this basis by normalizing each of the columns to have range 1. This ensures that the coefficients will have a limited range which allows us to specify a single prior variance and proposal mechanism for all coefficients. Note that when creating predicted values, the basis must be normalized using the range for the original design. The calculation of the basis is implemented in $C$ for computational speed.

**Acknowledgments**

I wish to thank Michael Newton for helpful discussions about this work.

**References**


