Splines Nonparametric Regression via reversible jump MCMC

Ronaldo Dias

Universidade Estadual de Campinas.

E-mail address: dias@ime.unicamp.br

Carolina C. Gomes

Universidade Estadual de Campinas.

Abstract

A standard procedure in nonparametric regression is to estimate the regression curve by using a linear combination of B-splines. In this context, the coefficients of the expansion, the number of knots and the knots' positions are unknown. The procedure to obtain the estimate of the regression curve is the reversible jump MCMC (Green, 1995).

1 Introduction

Since the pioneer work of Craven and Wahba (1979) using splines, several methods for nonparametric regression were suggested. Recently, Luo and Wahba (1997) and

Dias (1999) proposed more adpative methods to obtain good estimates of the regression curves. Under the Bayesian scheme, Denilson, Mallick and Smith (1998) suggest selecting regression models by using reversible jump MCMC (Green, 1995). Different from other methods this work shows a more general approach by also considering prior information on the knots positions and the coefficients of the expansions.

Suppose we have the following regression model,

$$y_i = f(x_i) + \varepsilon_i \quad i = 1, \ldots, n.$$

where ε_i's are uncorrelated with a $N(0, \sigma^2)$. Moreover, assume that the parametric form of the regression curve f is unknown. An attempt to solve this problem a nonparametric regression model is suggested by supposing that $f(x_i) = \sum_{j=1}^{K} \beta_j B_j(x_i, t)$, where the B_j's are the splines basis functions with a knot sequence $t = (t_1, \ldots, t_k)$. In particular the basis functions used here is,

$$B_j(x, t) = \beta_{j,0} + \beta_{j,1} t + \beta_{j,2} t^2 \beta_{j,3} t^3 + \sum_{i=1}^{k} \beta_{j,i} (x - t_i)_+^3,$$

with $(x)_+^3 = \max(0, x^3)$. Hence the problem becomes to estimate the vectors β, t, the dimension space K and σ.

2 Bayesian Approach

The Bayesian approach is to let π be the posterior density of the parameters (β, t, K, σ^2) given the vector $y = (y_1, \ldots, y_n)$, that is:

$$\pi(\beta, t, K, \sigma^2) \propto l_y(\beta, t, K, \sigma^2) p(\beta, t, K, \sigma^2),$$

where is the likelihood $l_y(\beta, t, K, \sigma^2) = \prod_{i=1}^{n} f(y_i | \beta, t, K, \sigma^2)$. Let's assume that

$$p(\beta, t, K, \sigma^2) = p(\beta | K, t) p(t | k) p(K) p(\sigma^2)$$

and the priors,

1. $K \sim Poisson(\lambda)$ (truncated) where the possible values of K is $1, \ldots, k_{max}$ such that $k_{max} < n$ to avoid interpolation.

2. $(t|K)$ the order statistics of k points from $\mathcal{U}(x_{(1)}, x_{(n)})$ with $x_{(i)}$ the i-th order statistics.

3. $(\beta|t, K) \sim N_K(\beta_0, \Sigma)$

4. $\sigma^{-2} \sim Gamma(a, b)$

By applying reversible jump MCMC (Green, 1995) we have four moves. The first move is the move of the vector of coefficients β. Choose randomly, β_j among $\beta_1, \ldots, \beta_K$ and let $\beta'_j = \beta_j + N(0, c^2)$.

The correspondent acceptance probability for this move is:

$$\min\left\{\frac{l_y(\beta', t, K, \sigma^2)p(\beta'|K, t)p(t|k)p(K)p(\sigma^2)}{l_y(\beta, t, K, \sigma^2)p(\beta|K, t)p(t|k)p(K)p(\sigma^2)}, 1\right\}$$

The second move is on the knots positions. Choose at random t_j among $t_1, \ldots, t_k$. Define $t_0 = 0$ and $t_k = 1$. Then choose $t'_j \sim U(t_{j-1}, t_{j+1})$.

Then acceptance probability is,

$$\min\left\{\frac{l_y(\beta, t, K, \sigma^2)p(\beta|K, t')p(t'|k)p(K)p(\sigma^2)}{l_y(\beta, t, K, \sigma^2)p(\beta|K, t)p(t|k)p(K)p(\sigma^2)}, 1\right\}$$

The third movement is the birth of a knot. For this, choose t^* for the proposed new knot position uniformly distributed on $[t_1, t_n]$. This must lie with probability 1 within an existing interval, say, (t_j, t_{j+1}). The new set of knots is given by:
$t'_1 = t_1, \ldots, t'_j = t_j, t'_{j+1} = t^*, t'_{j+2} = t_{j+1}, \ldots, t'_{k+1} = t_k$.

For this particular choice of basis functions each knot is associate to a single correspondent basis function. Take $\beta = (\psi, \beta)$, then $\psi'_j = \psi_j$, $j = 0, \ldots, 3$ and $\beta'_1 = \beta_1, \ldots, \beta'_j = \beta_j, \beta'_{j+1} = \beta^*, \beta'_{j+2} = \beta_{j+1}, \ldots, \beta'_{k+1} = \beta_k$ with β^* chosen from $N(\beta_{0j}, \sigma_j^2)$. Observe that this transformation is one-to-one and its Jacobian is 1. Following Green (1995), the acceptance probability for the birth of a knot is,

$$\alpha = \min\{1, (\text{likelihood}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian})\}.$$

The proposal ratio is,

$$\frac{(1/(k+1)) \times d_{k+1}}{(1/L) \times b_k \times \frac{1}{\sqrt{2\pi}\sigma_j} \exp(\beta^* - \beta_{0j})^2/2\sigma_j^2},$$

with $b_k = C \min\{1, \frac{p(k+1)}{p(k)}\}$, $d_{k+1} = C \min\{1, \frac{p(k)}{p(k+1)}\}$ and L is the range of the independent variable x. Choose C as large as possible subject to $b_k + d_{k+1} \leq 0.9$ for all $k = 1, \ldots, k_{max}$. This choice of C is to ensure that $b_k p(k) = d_{k+1} p(k+1)$. The prior ratio is given by,

$$\frac{p(\beta^*, t^*, k+1)}{p(\beta, t, k)} = \frac{p(\beta^*|k+1)p(t^*|k+1)p(k+1)}{p(\beta|k)p(t|k)p(k)},$$

with

$$\begin{aligned}
\frac{p(\beta^*|k+1)}{p(\beta|k)} &= \left(\frac{k}{k+1}\right)^{k/2} \left(\frac{k+1}{c}\right)^{1/2} \\
&= \times \exp\Big\{ -\frac{(k+1)^2}{c^2} \sum_{j=1}^{k+1} (\beta_j^* - \beta_{0j})^2 \\
&= + \frac{k^2}{c^2} \sum_{j=1}^{k} (\beta_j - \beta_{0j})^2 \Big\}
\end{aligned}$$

and

$$\frac{p(t^*|k+1)p(k+1)}{p(t|k)p(k)} = \frac{\lambda}{(k+1)^2}.$$

Since each knot is associated to a single basis function the likelihood ratio becomes,

$$lr_{\beta,t} = \exp\{\sum_{i=1}^{n} \beta^* B(x_i, t^*)\}$$

Simlilarly to the birth of a knot the fourth move takes in consideration to remove a single knot. Then new knot sequence is $t'_1 = t_1, \ldots, t'_{j-1} = t_{j-1}, t'_j = t^{j+1}, t'_{j+1} = t_{j+2}, \ldots, t'_{k-1} = t_k$, and the acceptance probability is built accordingly to the third movement.

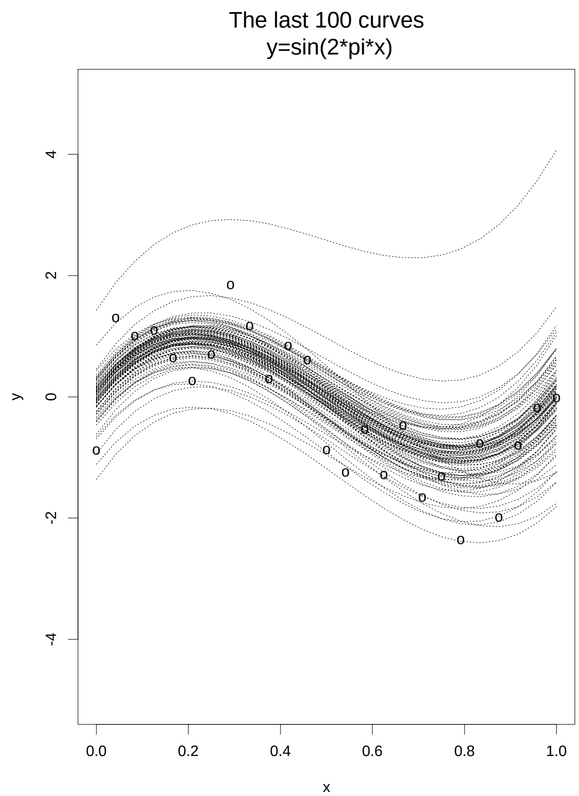

Figure 2.1: The last one hundred curve estimates

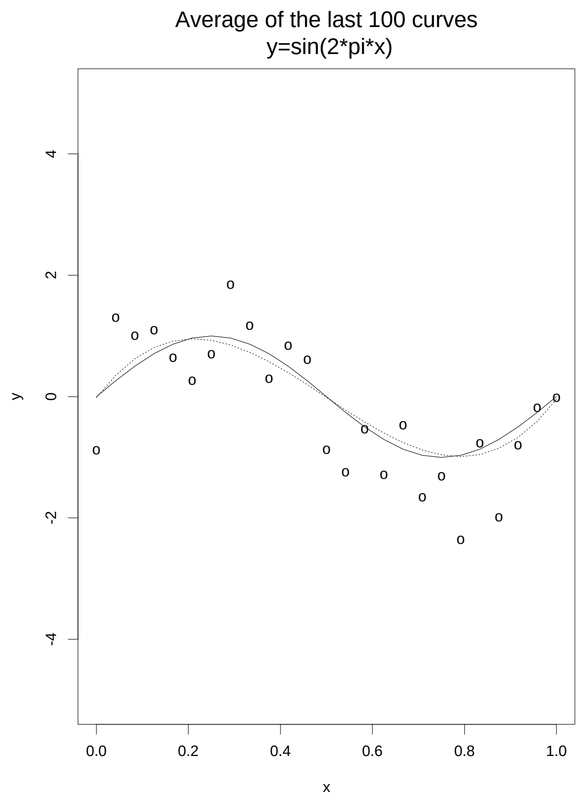

Figure 2.2: The last one hundred curve estimates

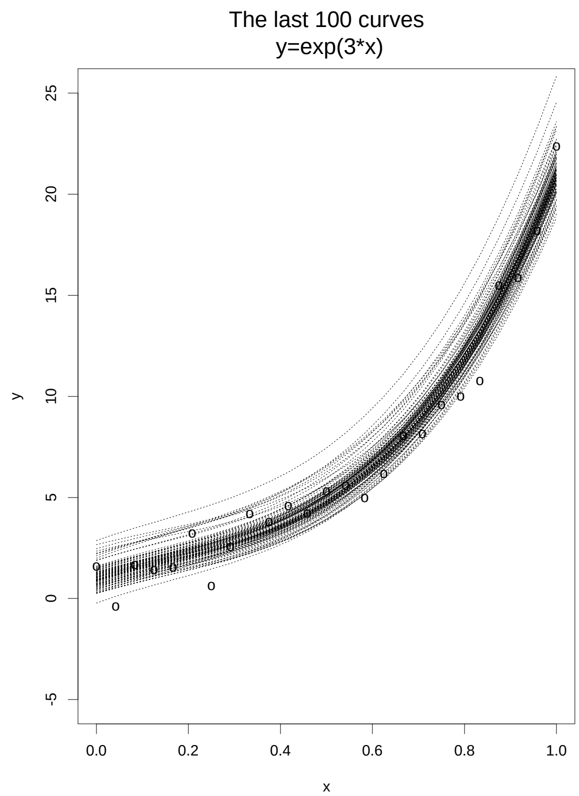

Figure 2.3: The last one hundred curve estimates

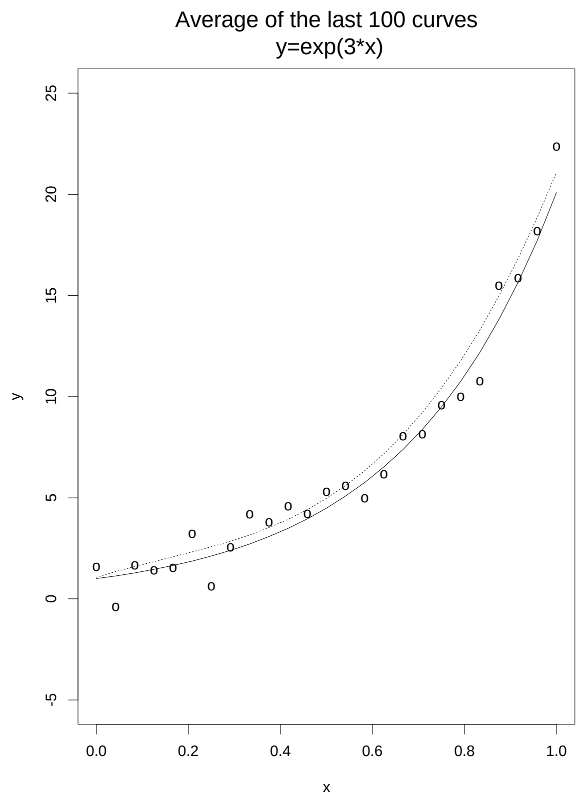

Figure 2.4: The last one hundred curve estimates

References

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numerische Mathematik* **31**: 377–403.

Denilson, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998). Automatic bayesian curve fitting, *Journal of the Royal Statistical Society B* **60**: 363–377.

Dias, R. (1999). Sequential adaptive non parametric regression via h-splines, *Communications in Statistics: Computations and Simulations* **28**: issue 2.

Green, P. J. (1995). Reversible jump Markov Chain Monte Carlo computation and bayesian model determination, *Biometrika* **82**: 711–732.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines, *Journal of the American Statistical Association* **92**: 107–116.