Approaches to Bayesian Smooth Unimodal Regression

George Woodworth

Dec 13, 1999

(Draft - please send comments to george-woodworth@uiowa.edu)

1. Background

Speech Articulation Data

The data in Figure 1 were obtained by asking 51 subjects ranging in age from 8 to 73 years to track, using jaw movements alone, a dot moving sinusoidally in one dimension at .6 Hz on a video monitor. Their jaw movements were captured electronically by means of strain guages and translated into the motion of a cursor which was to track (follow) the moving dot. Fidelity of tracking was measured in several ways, including TTD, the RMS difference between target and tracker shown in Figure 1. Low values of TTD indicate that the subject has high control of the speech articulator ((lips, jaw, or voice). The investigators believed ability to control speech articulators to be unimodal in age, reaching a broad optimum in early to middle adulthood. Hence, they wished to fit unimodal regressions to such data. The purpose was to establish age-norms against which to compare the performance of neurologically compromised individuals.

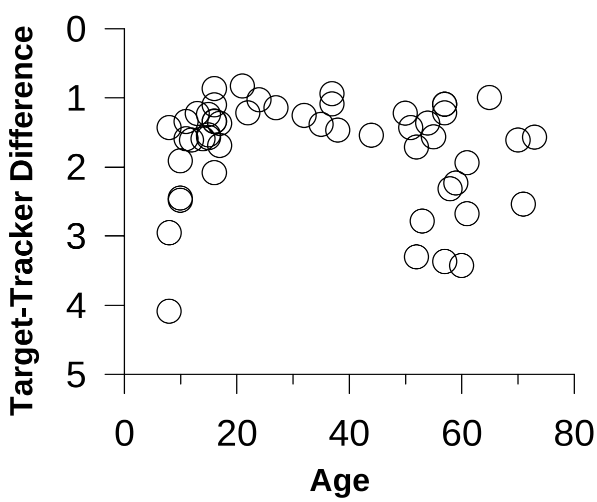

Figure 1. Jaw RMS target-tracker difference
for a .9Hz sinusoidal target.

Unimodal Regression

Unimodal regression is a type of order restricted inference (ORI). In the most common formulation of ORI, y, a px1 vector of homoskedastic, independent, normally distributed random variables, has mean vector μ which is know a priori to lie in a cone C. I.e. if μ and ν are in C then μa +νb is also in C for any non-negative scalars a and b. The set of unimodal vectors is not a cone unless the position of the mode is known and therefore the most powerful tools of ORI are not available in this case. To work around this difficulty, Frisen (1986), suggested estimating μ for all possible positions of the mode and selecting the mode with the smallest residual sum of squares. He suggested that when a smooth estimate is required the initial estimate could be smoothed with a unimodal, non-negative kernel, which preserves unimodality under convolution. Convex functions are either unimodal or monotone and lie in a cone; consequently, the problem of covex (or concave) regression has received more attention. However, convexity (or concavity) is considerably more restrictive than unimodality.

In this paper, I propose a Bayesian approach to smooth unimodal regression.

2. Bayesian, Smooth, Unimodal Regression

Let $y_j, 1 \le j \le n$, be independent observations from n subjects. Let $x_1, \ldots, x_n$ be values of a covariate, e.g. age. y_j is normally distributed with mean $\mu\left(x_j\right)$ and precision τ. The mean function $\mu\left(\cdot\right)$ is assumed a priori to be smooth; i.e., it has a continuous second derivative. Let $t_1, \ldots, t_p,\ p \le n,$ be the distinct covariate values. Sufficient statistics are $n_i,\ \bar{y}_i,\ and\ \ ss_i = \sum_{x_j = t_i} (y_j - \bar{y}_i)^2$, and the log likelihood function is

$$l(\mu,\tau) = \frac{n}{2}\ln(\tau) - \frac{\tau}{2}\left(\sum_{i=1}^{p} n_i \left(\bar{y}_i - \mu\left(t_i\right)\right)^2 + ss_i \right) \tag{2.1}$$

Bayesian analysis requires the specification of a prior distribution for $\tau > 0$ and a prior condtitional distribution for $\mu(\cdot) \mid \tau$ over the space of smooth, unimodal functions (or some subset of that space).

Sets of Smooth, Unimodal Functions

Let $u(t)$ be a smooth unimodal function over the interval [0,1]. If the mode is an endpoint, then $u(t)$ is monotone, otherwise there will be an interior mode $u_0 = u(t_0)$ at all points in the modal interval $\underline{t} \le t_0 \le \overline{t}$. If u is strictly unimodal, then $\underline{t} = t_0 = \overline{t}$.

Define

$$m(t) = m_0 + \text{sign}(t - t_0)\sqrt{(u_0 - u(t))} \tag{2.2}$$

Clearly $m(t)$ is non-decreasing and is smooth for $t \notin [\underline{t}, \overline{t}]$, and has at least one continuous derivative at $\underline{t}$ and $\overline{t}$. Thus if m(t) is any smooth, monotone function, then the function

$$v(t) = u_0 - (m_0 - m(t))^2 \tag{2.3}$$

is smooth and unimodal; consequently, functions of form (1.3) are a subset of smooth, unimodal functions. An example of a smooth, unimodal function which cannot be expressed this way is $v(t) = -|t|^3$.

Representation (2.3) reduces the problem to estimation of a smooth monotone function, which has been extensively investigated. Ramsay (1998) proved that any smooth monotone function has the representation

$$m(t) = c_0 + c_1 \int_0^t \exp\left(\int_0^s w(u)du\right)ds, \tag{2.4}$$

where $w(u)$ is any square integrable function. He proposed a penalized maximum likelihood estimation based on the likelihood function,

$$l(\beta, \tau, w) = -\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 m(t_i))^2 - \frac{\lambda}{2}\int_0^1 w^2(t)dt \tag{2.5}$$

The Bayesian interpretation of the penalty is that w=dW, where W is a Wiener process with precision λ. Combining (2.3) and (2.4), I propose the prior specification,

$$u(t) = u_0 + \left(u_1 - u_2\int_0^t \exp\left(\lambda^{-.5}\int_0^s dW(u)\right)ds\right)^2, \tag{2.6}$$

where $W(\cdot)$ is a standard Wiener process.

The smooth monotone function $m(t)=\int_0^t \exp(\lambda^{-.5}\int_0^s dW(u))ds$ evaluated at the points $t_1,\ldots,t_p$ has the nonlinear state space representation,

$$\begin{aligned} m_{i+1} &= m_i + m_i' e^{u_i} \\ m_{i+1}' &= m_{i+1}' e^{v_i} \end{aligned}, \tag{2.7}$$

where $(u_{i+1},v_{i+1}), 1\le i\le p-1$, are independent random vectors defined by,

$$\begin{aligned} u_{i+1} &= \ln\left(\int_{t_i}^{t_{i+1}} \exp(\lambda^{-.5}W(s))ds\right). \\ v_{i+1} &= W(t_{i+1}) - W(t_i) \end{aligned} \tag{2.8}$$

Although this approach is promising, its implementation requires a good approximation to the joint distribution of (u_{i+1},v_{i+1}). For that reason, I propose a second approach using b-splines.

3. B-spline representation of a smooth unimodal function.

The normalized cubic b-spline basis functions are smooth $b_{-1}(t),\ldots,b_{K+2}(t)$ with knots at $T_1,\ldots,T_k$ have the property that the function

$$v(t)=\sum_{j=-1}^{K+2} v_j b_j(t), \tag{3.1}$$

has no more strong sign changes than does the finite weight sequence $v_{-1},\ldots,v_{K+2}$ (Schumaker1981, Theorem 4.76). Since the basis functions sum to one for all t, this implies that if the finite series is $v_0,\ldots,v_{K+1}$ is unimodal, then the smooth function $v(t)$ is either unimodal or monotone. Thus functions of the form (3.1) with unimodal weight vectors $v_{-1},\ldots,v_{K+2}$ are a subset of smooth monotone functions.

For a Bayesian analysis, the data precision, τ, can be given a conjugate (gamma) prior $h(\tau)$ The problem is to specify a prior distribution $g(\mathbf{v}\,|\,\tau)$ for the weights $v_{-1},\ldots,v_{K+2}$ over the set of unimodal vectors. Once the priors are specified, the log posterior distribution, up to an additive function of the data, is,

$$l(\beta,\tau,\mathbf{v}) = -\frac{\tau}{2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 u(t_i))^2 + g(\mathbf{v}) + h(\tau), \tag{3.2}$$

where $$u(t) = \sum_{j=-1}^{K+2} v_j b_j(t) \qquad (3.3)$$

To specify the prior distribution of $v_{-1},\ldots,v_{K+2}$, note that the vector $v_{-1},\ldots,v_{K+2}$ is unimodal if and only if it can be expressed in the form $v_j = v_* + (m_* - m_j)^2$, where $m_{-1} \le m_0 \le \cdots \le m_{K+2}$ is non-decreasing. There are several possibilities for specifying a prior distribution over the space of monotone vectors; order statistics of independent and identically distributed random variables, cumulative sums of independent, non-negative random variables, etc. Informal investigation suggests that if the prior is fairly diffuse, the exact specification is not critical; however, this point requires further work.

Figure 2 shows the posterior mean unimodal b-spline curve with knots at 16,24,32,...,80. Data precision was given a diffuse gamma prior with mean 50 and shape parameter 0.5, which means that the residual standard deviation had prior median 0.21 and with prior probability .95 was between .06 and 4.5. The prior distribution of the unimodal weights was specified by,

$$v_j = v_* + (v_*^{.5} - m_j)^2, \quad m_j = |z_1| + \cdots + |z_j|, \qquad (3.4)$$

where $z_0,\ldots,z_{K+1}$, are independent, normally distributed random variables with zero mean and precision 1.0, and v_* has a Normal distribution with mean 1 and precision .01 truncated at zero. The posterior mean is shown in Figure 2; the posterior distribution of the residual standard deviation had mean 0.62 and 95% posterior credible interval .48 to .80. The posterior distribution was not sensitive to the hyperparameters (the mean and shape of the gamma distribution of τ, the precision of $z_0,\ldots,z_{K+1}$, and the mean and precision of v_*). Computations were carried out via Markov chain Monte-Carlo (Gelman, et al., 1996; Gilks, et al. 1996) using WinBUGS 1.2 (Spiegelhalter, et al., 1999).

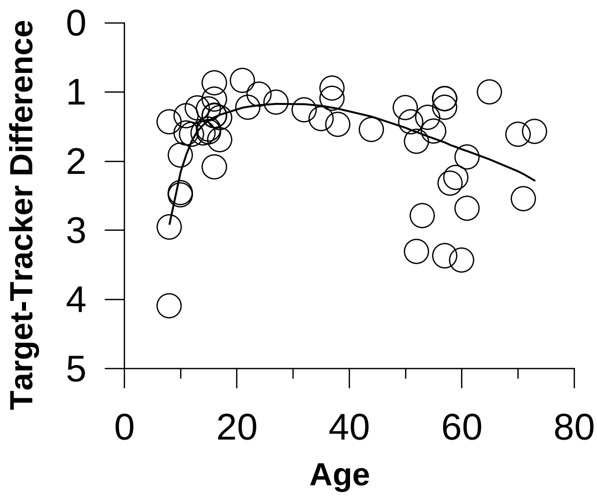

Figure 2. Unimodal b-spline fit.

References

Frisen, M. "Unimodal Regression," *The Statistician*, 35, 479-485, 1986.

Ramsey, J.O. "Estimating Smooth Monotone Functions," *J.R.Statisti.Soc. B*, 60, Part 2, 375-375, 1998

Schumaker, L.L. *Spline Functions: Basic Theory*, New York, John Wiley & Sons, 1981.

Gelman A. and Rubin DB. Markov chain Monte Carlo methods in biostatistics. Statistical Methods in Medical Research 1996; 5:339-355.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. *Markov Chain Monte Carlo in Practice*, London, Chapman and Hall, 1996.

Spiegelhalter DJ, Thomas A, Best NG. WinBUGS Version 1.2 User Manual. MRC Biostatistics Unit, Cambridge, UK, 1999.