HHS Public Access

Author manuscript

Biometrika. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

Biometrika. 2017 December ; 104(4): 939–952.

Bayesian Local Extremum Splines

M. W. WHEELER,

National Institute for Occupational Safety and Health, 1150 Tusculum Avenue, Cincinnati, Ohio 45226, MS C-15

D. B. DUNSON, and

Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708

A. H. HERRING

Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708

Summary

We consider shape restricted nonparametric regression on a closed set $\mathscr{X} \subset \mathbb{R}$, where it is reasonable to assume the function has no more than H local extrema interior to $\mathscr{X}$. Following a Bayesian approach we develop a nonparametric prior over a novel class of local extremum splines. This approach is shown to be consistent when modeling any continuously differentiable function within the class considered, and is used to develop methods for testing hypotheses on the shape of the curve. Sampling algorithms are developed, and the method is applied in simulation studies and data examples where the shape of the curve is of interest.

Keywords

Constrained function estimation; Isotonic regression; Monotone splines; Nonparametric; Shape constraint

1. Introduction

This paper considers Bayesian modeling of an unknown function $f_0 : \mathscr{X} \to \mathbb{R}$, where it is known that f_0 has at most H local extrema, or change points, interior to $\mathscr{X}$, and one wishes to estimate the function subject to constraints or test the hypothesis the function has a specific shape. For example, one may wish to consider a monotone function versus one having an N shape. We propose a spline construction that allows for nonparametric estimation of shape-constrained functions having at most H change points. The approach places a prior over a knot set dense in $\mathscr{X}$, and, to sample over the models defined by this knot set, a Markov chain Monte Carlo algorithm is developed to sample models. The method allows for nonparametric hypothesis testing of different shapes within the class of functions considered.

The shape-constrained regression literature focuses primarily on functions that are monotone, convex, or have a single minimum; that is, cases with $H \leq 1$. Ramgopal et al. (1993), Lavine & Mockus (1995), and Bornkamp & Ickstadt (2009) consider priors over cumulative distribution functions used to model monotone curves. Holmes & Mallick (2003), Neelon & Dunson (2004), Meyer (2008), and Shively et al. (2009) develop spline-

based approaches for monotone functions. Hans & Dunson (2005) design a prior for umbrella-shaped functions, while Shively et al. (2011) propose methods for fixed- and free-knot splines that model continuous segments having a single unknown change point.

Extending these approaches to broader shape constraints is not straightforward. For example, to obtain $H = 3$ change points, one could define a prior over B-spline bases de Boor (2001, page 87)) having four monotone segments that alternately increase and decrease. However, for even a moderate number of pre-specified knots and a known number of change points, allowing for uncertainty in the locations of the change points leads to a daunting computational problem. Bayesian computation via Markov chain Monte Carlo is subject to slow mixing and convergence rates in alternating between updating the spline coefficients conditionally on the change points and vice versa, and it is not clear how to devise algorithms that can efficiently update both simultaneously. These difficulties are compounded by allowing for the possibility that some of the change points should be removed, which is commonly the situation in applications. By defining a new spline basis based on the number of change points, we bypass these issues.

Little work has been done on nonparametric Bayesian testing of curve shapes. Salomond (2014) and Scott et al. (2015) consider Bayesian nonparametric testing for monotonic versus an unspecified nonparametric alternative, but do not consider shapes beyond monotonicity. Our approach allows for testing of all shapes, where shape is defined as the type and sequence of extrema. For example, one can use this approach to test for an umbrella shape verses an N-shaped curve and use the same procedure to test the umbrella shape against monotone alternatives.

We propose a new approach to incorporating shape constraints based on splines that are carefully constructed to induce curves having a particular number of extrema. This is similar in spirit to the I-spline construction of Ramsay (1988) or the C-spline construction for convex splines (Meyer, 2008; Meyer et al., 2011), both of which create a spline construction based upon the derivative of the spline. When paired with positivity constraints on the spline coefficients, our construction enforces shape restrictions on the curve of interest by limiting the number of change points.

Another key aspect of our approach is that we place a prior over a countable dense set of knots, which allows the number of the splines in the model space to grow. This bypasses the sensitivity to choice of the number of knots, while facilitating computation and theory on consistency. In particular, we propose a prior over nested model spaces where the location of the knots is known for each model. This allows for a straightforward reversible jump Markov chain Monte Carlo algorithm (Green, 1995) based upon Godsill (2001). This is different from much of the previous Bayes literature allowing unknown numbers of knots (Biller, 2000; DiMatteo et al., 2001). In these methods, the knot locations are unknown, and the reversible jump Markov chain Monte Carlo proposal must propose a knot to add or delete as well as its location. Such algorithms are notoriously inefficient.

2. Model

2·1. Local Extremum Spline Construction

Let $\mathcal{F}^H$ be a set of functions defined on the closed set $\mathcal{X} \subset \mathbb{R}$, such that for $f_0 \in \mathcal{F}^H$, f_0 is continuously differentiable and has H or fewer local extrema interior to $\mathcal{X}$. Such functions can be modeled using B-spline approximations of the form

$$f(x) = \sum_{k=1}^{K+j-1} \beta_k B_{(j,k)}(x). \quad (1)$$

Here, β_k is a scalar coefficient and $B_{(j,k)}(x)$ is a B-Spline function of order j defined on the knot set $\mathcal{T} = \{\tau_k\}_{k=1}^{K}$, $\tau_1 \leq \ldots \leq \tau_K$, which includes end knots. For any knot set, de Boor (2001, page 145) showed that there exist spline approximations such that $\|f - f_0\|_\infty \leq \Delta \|f_0\|_\infty$, where Δ is the maximum difference between adjacent knots. Though this construction can be used to model f_0 with arbitrary accuracy, it does not ensure that the approximating function f is itself in $\mathcal{F}^H$.

We force $f \in \mathcal{F}^H$ to have at most H local extrema by defining a new spline basis

$$B^*_{(j,k)}(x) = M \int_{-\infty}^{x} \left\{ \prod_{h=1}^{H} (\xi - \alpha_h) \right\} B_{(j,k)}(\xi)\, d\xi, \quad (2)$$

where $B_{(j,k)}(x)$ is a B-spline constructed using the knot set $\mathcal{T}$, $\{\alpha_1, \ldots, \alpha_h\}$ are distinct change points, and M is a fixed integer. Letting $B^*_{(j,0)}(x) = 1$, if $\beta_k \geq 0$, for all $k \geq 1$, any linear combination of local extremum spline basis functions for any distinct values of $\alpha_1, \ldots, \alpha_H$ in (2) will be in $\mathcal{F}^H$.

Proposition 1—If $f(x) = \sum_{k=0}^{K+j-1} \beta_k B^*_{(j,k)}(x)$ for any $K \geq 1$ with $M \in \{-1, 1\}$, $j \geq 1$, and $\beta_k \geq 0$ for all $k \geq 1$, then $f \in \mathcal{F}^H$.

This result follows from the constraint on the β_k coefficients. By forcing $\beta_k \geq 0$ for $k \geq 1$, the sign of the derivative is controlled by the polynomial $M \prod_{h=1}^{H} (x - \alpha_h)$, which allows a maximum of H local extrema located at the change points $\{\alpha_1, \ldots, \alpha_H\}$. When $\beta_k = \ldots = \beta_{k+1} = 0$ and $\alpha_h \in [\tau_{k+j}, \tau_{k+j+1}]$, α_h does not define a unique extremum. In this case, there is a flat region, and multiple configurations of the change point parameters can give the same curve. Otherwise, the extrema are uniquely defined for all $\alpha_h \in \mathcal{X}$, and fewer than H extrema can be considered if $\alpha_h \notin \mathcal{X}$.

Theorem 1—For any $f_0 \in \mathcal{F}^H$ *and* $\varepsilon > 0$ there exist a knot set $\mathcal{T}$ and a local extremum spline f^{LX} defined on this knot set such that

$$\|f_0 - f^{LX}\|_\infty < \varepsilon.$$

The flexibility of local extremum splines is attributable to the B-splines used in their construction. The proof of Theorem 1 assumes that M can be chosen to be positive or negative, which allows all functions in $\mathcal{F}^H$ to be approximated. If M is fixed, then any function with $H - 1$ extrema can be modeled. For exactly H extrema, the approach is limited to modeling functions that are either initially increasing or initially decreasing, and this depends on the sign of M.

Though the polynomial weighting does not affect the ability of the local extremum spline to model arbitrary functions in $\mathcal{F}^H$, it does impact the magnitude of the spline, $sup_{x \in \mathcal{X}} |B^*_{(j,k)}(x)|$, which may cause difficulty in the prior specification. To minimize this effect it is often beneficial to construct the splines on the interval (−0·5, 0·5). Additionally, it is often beneficial to multiply M by a fixed constant to aid in prior specification.

2·2. Infill Process Prior

Bayesian methods for automatic knot selection (Biller, 2000; DiMatteo et al., 2001) commonly define priors over the number and location of knots. Using free knots presents computational challenges, while fixed knots are too inflexible; we address this by defining a prior over a branching process where the children of each generation represent knot locations that are binary infills of the previous generation. This defines a nested set of spline models such that successive generations produce knots that can be arbitrarily close.

To make these ideas explicit, define $\mathcal{T}_N = \left\{ a/2^{N+1} : a = 1, 3, \ldots, 2^{N+1} - 1 \right\}$ with $N \in \{0, 1, 2, 3, \ldots\}$. Assume for the sake of exposition, and consider an infinite complete binary tree. In this tree, each node at a given depth N is uniquely labeled using an element from $\mathcal{T}_N$. If the node's label is $a/2^{N+1}$, its children are labeled $(2a - 1)/2^{N+2}$ and $(2a + 1)/2^{N+2}$. For example, the node labeled 3/8 at $N + 2$ has children labeled 5/16 and 7/16, and the root node labeled 1/2 has children labeled 1/4 and 3/4.

We induce a prior on the set of local extremum spline basis functions through a branching process over this tree. The process starts at the root node $N = 0$ where the generation of children occurs via two independent Bernoulli experiments having probability of success ζ. On each success, a child is generated, and its label is added to the knot set. This process repeats until it dies out. If $\zeta < 0{\cdot}5$, the probability of extinction is 1 (Feller, 1974, page 297). To favor parsimony, we define the probability of success for a node at a given depth N to be $0{\cdot}5^{N+1}$, which decreases the probability of adding a new node the larger the tree becomes. The tree $\mathcal{M}$ generated from this process corresponds to a knot set $\mathcal{T}_{\mathcal{M}}$. We complete the knot set by adding end knots {0, 1}.

Letting $K = |\mathcal{T}_{\mathcal{M}}|$ be the number of knots for tree $\mathcal{M}$ including end knots, there are $K + j - 1$ basis functions. Letting β_k denote the coefficient on $B^*_{(j,k)}(x)$, we choose the prior:

$$p(\beta_k|\mathscr{M}) = \pi 1_{(\beta_k=0)} + (1-\pi)\,\mathrm{Exp}(\beta_k;\lambda), \quad 1 \le k \le K+j-1, \tag{3}$$

where Exp(β_k; λ) is an exponential distribution with rate parameter λ, π is the prior probability of $\beta_k = 0$, and the β_k are drawn independently conditionally on $\mathscr{M}$, π, and λ. For the intercept, we let $\beta_0 \sim N(0, c)$, and we allow for greater adaptivity to the data through hyperpriors, $\pi \sim \mathrm{Be}(\nu, \omega)$ and $\lambda \sim \mathrm{Ga}(\delta, \kappa)1(\lambda > \varepsilon)$, which is a truncated gamma distribution, that is truncated slightly above zero to guarantee posterior consistency. In practice, this value is set to 10^{-5}, making the prior indistinguishable from the Gamma distribution.

To allow uncertainty in locations of the change points, we choose the prior

$$p(\alpha) = \prod_{h=1}^{H} \mathrm{TN}\{\alpha_h; (b-a)/2, 1, a, b\} \tag{4}$$

where TN{($b - a$)/2, 1, a, b} is a normal distribution with mean ($b - a$)/2 and variance 1, truncated below by a and above by b with $\mathscr{X} \subset [a, b]$. If $\alpha_h \le inf\,\mathscr{X}$ or $\alpha_h \ge sup\,\mathscr{X}$, then the change point is removed. We assume that M is pre-specified corresponding to prior knowledge of whether the function is initially increasing or decreasing, though generalizations to place a Bernoulli or alternative prior on M are straightforward.

The prior for the change point parameters is defined such that $\mathscr{X} \subset [a, b]$. A change point placed outside of $\mathscr{X}$ allows the derivative of f to be non-zero at inf $\mathscr{X}$ or sup $\mathscr{X}$. In practice, results are insensitive to the choice of a and b. In what follows, we choose $a = inf(\mathscr{X}) - \Delta$ and $b = sup(\mathscr{X}) + \Delta$, where $\Delta = \{sup(\mathscr{X}) - inf(\mathscr{X})\}/2$.

2·3. Prior Properties

Define $\mathscr{F}^{H+}$ as the space of continuously differentiable functions with H or fewer local extrema, such that, for all $f_0 \in \mathscr{F}^{H+}$ having exactly H extrema, the first extremum from the left is a maximum, and, for all functions in $f_0 \in \mathscr{F}^{H+}$ having less than H extrema, the function is also in $\mathscr{F}^{H-1}$. Conversely, define $\mathscr{F}^{H-}$ as the set of continuously differentiable functions with H or fewer local extrema, such that for all functions having exactly H extrema, the first from the left is a minimum, and for all functions $f_0 \in \mathscr{F}^{H-}$ having less than H extrema, they are also in $\mathscr{F}^{H-1}$. The prior places positivity in ε–neighborhoods of any f_0 in $\mathscr{F}^{H-}$ or $\mathscr{F}^{H+}$ depending on the sign of M.

Lemma 1—Letting f^{LX} be a randomly generated local extremum spline from the prior defined in §2·2 for all $f_0 \in \mathscr{F}^{H-1}$,

$$pr(\|f^{LX} - f_0\|_\infty < \varepsilon) > 0.$$

This holds for all $f_0 \in \mathcal{F}^{H+}$ if H is odd and $M < 0$ or H is even and $M > 0$. Otherwise, if H is even and $M > 0$ or H is odd and $M < 0$, this holds for all $f_0 \in \mathcal{F}^{H-}$.

Using this result we can show posterior consistency. Assume that $Y = (y_1, \ldots, y_n)^T$ are observed at locations $(x_1, \ldots, x_n)$ such that $y_i \sim \mathrm{N}\left\{f_0(x_i), \sigma_0^2\right\}$. Following Choi & Schervish (2007), assume that the design points are independent and identically distributed from some probability distribution Q on the interval $\mathcal{X}$, or observed using a fixed design such that $\max(|x_i - x_{i+1}|) < (K_1 n)^{-1}$, where $0 < K_1 < 1$ and $i < n$. Define the neighborhoods $W_{\varepsilon,n} = \{(f, \sigma) : \int |f(x) - f_0(x)| dQ_n(x) < \varepsilon, |\sigma/\sigma_0 - 1| < \varepsilon\}$ and $\mathrm{U}_\varepsilon = \{(f, \sigma) : d_Q(f, f_0) < \varepsilon, |\sigma/\sigma_0 - 1| < \varepsilon\}$, where $d_Q(f_1, f_2) = \inf \{\varepsilon > 0 : Q[\{x : |f_1(x) - f_2(x)| > \varepsilon\}] < \varepsilon\}$. Under the assumption that the prior over σ assigns positive probability to every ε–neighborhood of σ_0, one has:

Theorem 2—Let f^{LX} be a randomly generated curve from the prior defined in §2·2 with $f_0 \in \mathcal{F}^{H-1}$. If P_{f_0,σ_0} is the joint distribution of $\{y_i\}_{i=1}^{\infty}$ conditionally on $\{x_i\}_{i=1}^{\infty}$, $\{Z_i\}_{i=1}^{\infty}$ is a sequence of open subsets in $\mathcal{F}^{H-1}$ that is defined by $W_{\varepsilon,n}$ for fixed designs or by U_ε for random designs, and $\prod_n$ is the posterior distribution of f_0 given $\{y_i\}_{i=1}^{n}$, then

$$\prod\nolimits_n (f \in \mathcal{Z}_n^C | y_1, \ldots, y_n) \to 0 \ \text{ almost surely } \ [P_{f_0,\sigma_0}].$$

Further, for all H odd if $M < 0$, this relation holds for $f_0 \in \mathcal{F}^{H+}$, otherwise it holds for $f_0 \in \mathcal{F}^{H-}$. Similarly, for H even if $M > 0$, then $f_0 \in \mathcal{F}^{H+}$, otherwise it holds for $f_0 \in \mathcal{F}^{H-}$.

The proof of this consistency result follows from Choi & Schervish (2007) and the prior positivity result above. The condition on the prior over σ^2 can be satisfied with an inverse-Gamma distribution.

2·4. Bayes Factors for Testing Curve Shapes

Our approach allows one to define the shape of the curve through the a vector and to place prior probability on a class of functions having a given shape, i.e the number and type of extrema in $\mathcal{X}$. When there are flat regions of f_0 the shape of the curve is not uniquely identifiable based upon the configuration of a, and hypothesis tests may be inconclusive. For an example of this, see the consistency arguments for monotone curve testing in Scott et al. (2015). In what follows, we assume that $|f_0'(x)| > 0$ at all points in $\mathcal{X}$ except within flat regions.

Let $\mathbb{H}_1$ and $\mathbb{H}_2$ denote two distinct and non-nested sets of a values, corresponding to distinct shapes. These sets are defined by the number of $\alpha_h \in \mathcal{X}$, the number of $\alpha_h \leq inf(\mathcal{X})$, and the number of $\alpha_h \geq sup(\mathcal{X})$. One can compute pr($Y|f_0 \in \mathbb{H}_1$) and pr($Y|f_0 \in \mathbb{H}_2$), with the corresponding Bayes factor between the two shapes being

$$BF_{12} = \frac{\mathrm{pr}(Y | f_0 \in \mathbb{H}_1)}{\mathrm{pr}(Y | f_0 \in \mathbb{H}_2)}. \quad (5)$$

This quantity is not available analytically, but can be estimated through posterior simulation by monitoring the α and β vectors.

Any two shapes falling within $\mathcal{F}^H$ can be compared using this approach. Alternatively, one may be interested in the hypothesis that f_0 is in a class of functions with at least K extrema. For example, one may wish to assess whether or not the function is monotone. In this case, one can define $\mathbb{H}_1$ to correspond to functions in $\mathcal{F}^H$ with F or more extrema and $\mathbb{H}_2=\mathbb{H}_1^c$ to functions with less than F extrema. The value of H can be elicited as an upper bound on the number of extrema to avoid highly irregular functions. For such tests, the following result holds.

Proposition 2—Let $\mathbb{H}_1$ be the class of functions in $\mathcal{F}^H$ with F or more extrema and $\mathbb{H}_2=\mathbb{H}_1^c \cap \mathcal{F}^H$. If $f_0 \in \mathbb{H}_1$, then

$$BF_{12} \to \infty$$

as $n \to \infty$

This result, an application of Theorem 1 in Walker et al. (2004), It follows from the fact that local extremum spline representations having fewer than F change points can never be arbitrarily close to the function of interest.

3. Posterior Computation

We rely on Godsill (2001) to develop a reversible jump Markov chain Monte Carlo algorithm to sample between models. Consider moves between models $\mathcal{M}$ and $\mathcal{M}'$, where the model $\mathcal{M}'$ has one extra knot that is a child of a node also in $\mathcal{M}$. As described further in the Supplementary Material, most of the local extremum spline basis functions for model $\mathcal{M}$ and $\mathcal{M}'$ are identical, with only $j+2$ different functions. Let $\beta_{-\mathcal{M}}$ denote the coefficients on all the splines that are the same as well as σ^2, π and λ, which are parameters shared between both models. The remaining spline coefficients are $\beta_{\mathcal{M}}$ and $\beta_{\mathcal{M}'}$ for models $\mathcal{M}$ and $\mathcal{M}'$, respectively. As in Godsill, given the shared vector $\beta_{-\mathcal{M}}$, we marginalize $\beta_{\mathcal{M}}$ and $\beta_{\mathcal{M}'}$ out of the posterior to compute $p(\mathcal{M}' \mid Y, \beta_{-\mathcal{M}})$ and $p(\mathcal{M} \mid Y, \beta_{-\mathcal{M}})$. This marginalization requires numerical integration of multivariate normal distributions, which is performed using Genz (1992) and Genz & Kwong (2000). The probability of a move between two models is determined by the ratio

$$h=\frac{q\left(\mathcal{M};\mathcal{M}'\right)\, p\left(\mathcal{M}'|Y,\beta_{-\mathcal{M}}\right)}{q\left(\mathcal{M}';\mathcal{M}\right)\, p\left(\mathcal{M}|Y,\beta_{-\mathcal{M}}\right)}, \quad (6)$$

where a knot insertion is made with probability min(1, h), a knot deletion is made with probability min(1, 1/h), and $q(\mathcal{M}; \mathcal{M}')$ is the transition probability between $\mathcal{M}$ and $\mathcal{M}'$.

All proposals are made between models that are nested and differ by only one knot. When the current model has no children we propose a knot insertion with unit probability. Otherwise, the proposal adds or deletes a knot with probability 1/2, and the inserted or deleted knot is chosen uniformly. For a knot insertion, as we are going from model $\mathcal{M}$ to $\mathcal{M}'$, the available knots are represented by all failures in the branching process that generated $\mathcal{M}$. A knot deletion going from model $\mathcal{M}'$ to $\mathcal{M}$ represents all of the nodes in the branching process that generated $\mathcal{M}'$ that do not have any children. All other parameters, including the spline coefficients, are sampled in Gibbs steps described in the supplement.

The posterior distribution is often multimodal, with the sampler getting stuck in a single mode, when widely different parameter values have relatively large support by the data, with low posterior density between these isolate modes. To increase the probability of jumps between modes, a parallel tempering algorithm (Geyer, 1991, 2011) is implemented.

4. Simulation

4·1. Simulation Specification

We investigate our approach through simulations for functions having 0, 1, or 2 local extrema interior to $\mathcal{X}$. For all simulations, we place a $Ga(1, 1)$ prior over σ. For the hyper prior on π, we let $\nu = 2$ and $\omega = 18$, which puts low prior probability on flat curves. Additionally, for the hyper prior over λ, we let $\delta = 0{\cdot}2$ and $\kappa = 2$, which favors smaller values of β. All local extremum splines were constructed using B-splines of order 2 with $M = 100$.

The Markov chain Monte Carlo algorithm was implemented in the R programming language with some subroutines written in C++ and is available from the first author. Depending on the complexity of the function, the algorithm took between 60 and 90 seconds per 50, 000 samples using one core of a 3·3 gigahertz Intel i7-5830k processor. Parallelizing the tempering algorithm on multiple cores may substantially reduce the computation time. Additional information on the convergence of the algorithm, as well as impact of the B-spline order used, is provided in the Supplementary Material.

4·2. Curve Fitting

We compare the local extremum spline approach to other nonparametric methods, including Bayesian P-splines (Lang & Brezger, 2004), a smoothing spline method described in Green & Silverman (1993), and a frequentist Gaussian process approach described in Chapter 5 of Shi & Choi (2011). We consider seven different curves with between 0 and 2 extrema and compare the fits of the other approaches and of a local extremum spline specified to have at most $H = 2$ change points. The following true curves are investigated:

$$f_1(x) = 10x^2, \qquad f_2(x) = 2 + 20\Phi\{(x - 0\cdot 5)/0\cdot 071\},$$
$$f_3(x) = 5cos(\pi x), \qquad f_4(x) = 10(x - 0\cdot 5)^2,$$
$$f_5(x) = -2\cdot 5 + 10exp\left\{-50(x - 0\cdot 35)^2\right\}, \qquad f_6(x) = 1 + 2\cdot 5sin\{2\pi(x + 8)\} + 10x,$$
$$f_7(x) = 5\min(2\pi x)/(x + 0\cdot 75)^3 - 2\cdot 5(x + 10\cdot 5).$$

We set $y_i = f_j(x_i) + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$. Functions f_1, f_2 and f_3 are monotone, f_4 and f_5 have one change point, and f_6 and f_7 have two change points. For each simulation, a total of 100 equidistant points were sampled in $\mathcal{X} = [0, 1]$. We consider $\sigma^2 = 1, 4$. For each simulation condition, 250 data sets were generated, fitted and compared using the mean squared error, $n^{-1}\sum_{i=1}^{n}\left\{\hat{f}(x_i) - f(x_i)\right\}^2$, for the local extremum spline, smoothing spline, Bayesian P-spline, and Gaussian process approaches.

For the local extrema approach, we collected 50,000 Markov chain Monte Carlo samples, with the first 10, 000 samples disregarded as burn-in. For the parallel tempering algorithm, we specify 12 parallel chains with $\{\kappa_1, \ldots, \kappa_{12}\}$ = {1/30, 1/24, 1/12, 1/9, 1/5, 1/3·5, 1/2, 1/1·7, 1/1·3, 1/1·2, 1/1·1, 1}, and monitor the target chain with $\kappa_{12} = 1$. The P-spline approach was defined using 30 equally-spaced knots, and the prior over the second-order random walk smoothing parameter was IG(1, 0·0005), distribution, which was one of the recommended choices in Lang & Brezger (2004).·In this approach, 25,000 posterior samples were taken, discarding the first 5,000 as burn in. For the smoothing spline method, the R function 'smooth.spline' was used. Finally, the Gaussian process approach used a frequentist implementation given in the R package 'GPFDA.'

Table 1 gives the integrated mean squared error of the various approaches. All numbers marked with an asterisk are significantly different from local extremum splines. The local extremum approach integrated mean square error is always smaller than the others, and in most cases it is significantly different at the 0·05 level. Generally, when there is a high signal-to-noise ratio, the methods perform similarly, but when the ratio decreases, specifically in flat regions, the local extremum approach was superior as it removed artifactual bumps from the estimate.

4·3. Hypothesis Testing

We perform a simulation experiment investigating the method's ability to correctly identify the shape of the response function for three sets of hypotheses. In the first case, the null hypothesis is the set of all functions with one or more extremum, and the alternative, $\mathbb{H}_1$, is the set of all monotone functions. In the second test, the null consists of all monotone functions, and the alternative, $\mathbb{H}_2$, is all functions with one or more extremum. Finally, for the third test the null hypothesis is the set of functions having at most one extrema, and the alternative, $\mathbb{H}_3$, is the set of functions with two extrema first having a local maximum followed by a local minimum. Functions are defined on $\mathcal{X} = [0, 1]$. The nine functions used in this simulation are:

$$\begin{array}{ll} & \mathbb{H}_1 \\ g_1(x) = 2 + 0 \cdot 5x + \Phi\left\{(x - 0 \cdot 5)/0 \cdot 071\right\}, & g_2(x) = 0 \cdot 5 sin\left\{2\pi(x + 8)\right\} + 4 \cdot 75x, \\ g_3(x) = 1 + 2 \cdot 25x, & \end{array}$$

$$g_4(x) = -4(x-0\cdot75)^2, \qquad g_5(x) = 1+2x-1\cdot56exp\left\{-50(x-0\cdot5)^2\right\},$$
$$g_6(x) = 15(x-0\cdot5)^3 1_{(x<0\cdot5)} + 0\cdot3(x-0\cdot5) - exp\{-250(x-0\cdot25)\},$$

$$g_7(x) = 0.85sin\{2\pi(x+8)\} + 4\cdot75x, \qquad g_8(x) = g_5(x),$$
$$g_9(x) = 5sin(2\pi x)/(x+0\cdot75)^3 - 2\cdot5(x+10\cdot5) + 2.$$

For the simulation, data are generated assuming $y_i = g_j(x_i) + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma^2)$ and $\sigma^2 = 1$. We consider sample sizes $n = 100$, 200, 300, and 400, with 50 data sets constructed where points are sampled evenly across $\mathcal{X}$, for each sample condition. The local extremum approach is as above except, but 150,000 posterior samples are taken with the first 10,000 disregarded as burn-in. For tests $\mathbb{H}_1$ and $\mathbb{H}_2$, the local extremum approach is compared with the Bayesian method of Salomond (2014) and the frequentist methods of Baraud et al. (2005) and Wang & Meyer (2011). For the method of Baraud et al. we use the test where $\ell_n = 25$, and for the method of Wang and Meyer we use $k = 4$ splines, which were the most powerful tests presented in the respective articles.

The Bayesian tests produce Bayes factors, while the frequentist tests have corresponding test statistics. We compare the methods based upon area under the receiver operating curve. For the simulation, the false positive rate was computed from the values of the test statistics for the other functions not in the test set. As a frequentist calibration of our Bayesian test, one can choose a threshold on the Bayes factor to control the type I error rate at a specified level based on an approximation to the distribution of the Bayes factor under the null hypothesis. We describe this approximation in the Supplementary Material.

Figure 1 shows the receiver operating curve for hypothesis $\mathbb{H}_1$. This shows that the local extremum approach is superior to the other three approaches across all false positive rates. Further, the estimated area under the receiver operating curve is 0·94, better than the approaches of Salomond at 0·86, Baraud at 0·77, and Wang and Meyer at·0·74. When looking at the impact of sample size on the tests, the power of the local extrema approach increases as the sample size increases, does so at a rate greater than competitors, and is similarly superior for hypothesis $\mathbb{H}_2$, data not shown.

For hypothesis $\mathbb{H}_3$, there is not an equivalent methodology in the literature, but the performance of our approach is excellent. The area under the receiver operator curve is 0·94. For the Bayes factor cut point of 6, Table 2 gives results across all simulation conditions.·Our test achieves high power for function g_7, even though it differs this function is only slightly different from g_3. Function g_8 is the same as g_5, this simulation gives evidence that the departure from monotonicity may be due to the pronounced U shape in the data and not necessarily because there are two extrema, which requires more data to conclude in favor of $\mathbb{H}_3$.

4·4. Seasonal Influenza and Pneumonia Death Rate

In temperate climates, the prevalence of influenza peaks in the winter months while dropping in the warmer months. Estimating this seasonal effect as well as departures from this effect, may be of interest when estimating the magnitude of an influenza epidemic. Here, we expect a peak in the winter months followed by a trough in the summer months. Parametric models for this pattern may not be adequate to model the observed phenomena, and smoothing approaches do not guarantee this pattern. We use local extremum splines, setting $H = 2$, to estimate this trend for Virginia, North Carolina and South Carolina for data collected by the Centers for Disease Control and Prevention National Center for Health Statistics Mortality surveillance branch.

Figure 2 plots the estimated mortality rates, estimated using an additive model defined by a quadratic trend representing a decrease in mortality over time, a seasonal component defined using local extremum spline, and a P-spline that represents departures from the overall trend. This seasonal component is different from the trend published by the Centers for Disease Control (Viboud et al., 2010), mainly due to the asymmetry in the local extrema approach during the winter months, which cannot be captured by a single sinusoidal function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the referees and associate editor for comments on earlier versions of this manuscript. This research was partially supported by a grant from the National Institute of Environmental Health Sciences of the United States National Institutes of Health.

Appendix 1

Proofs of results

Proof of Proposition 1

It is well known that $\sum_{k=1}^{K+j-1} \beta_k B_{(j,k)}(x)$ is continuous for $j \geq 1$ and for $j \geq 1$ and for all $x \in \mathcal{X}$. Further, $\prod_{h=1}^{H}(x - \alpha_h)$ is a polynomial; therefore, $\prod_{h=1}^{H}(x - \alpha_h)\sum_{k=1}^{K+j-1} \beta_k B_{(j,k)}(x)$ is continuous with anti-derivative $\sum_{k=0}^{K+j-1} \beta_k B^*_{(j,k)}(x)$.

If $\beta_k \geq 0$ for all $k \geq 1$, then $\sum_{k=0}^{K+j-1} \beta_k B_{(j,k)}(x) \geq 0$ for all $x \in \mathcal{X}$ and $f'(x) = \prod_{h=1}^{H}(x - \alpha_h)\sum_{k=1}^{K+j-1} \beta_k B_{(j,k)}(x)$ can only change sign when $x = \alpha_h$. Thus, there are at most H local extrema interior to $\mathcal{X}$, with $f \in \mathcal{F}^H$. □

Proof of Theorem 1

Consider $f_0 \in \mathcal{F}^H$, where f_0 has exactly H change-points. Functions with less than H change points can be modeled by removing the required change point parameters from $\mathcal{X}$ and continuing with the proof below.

Let f^{BS} be a taut B-spline approximation of f_0 of order $j+1$ defined on the knot set $\mathcal{T}$ having exactly H extrema such that

$$\|f_0 - f^{\mathrm{BS}}\|_\infty < \Delta C.$$

Here f^{BS} is defined on $\mathcal{T}$, where $\Delta = \max_k |\tau_k - \tau_{k+j}| < 1$. As f_0 and f^{BS} are continuous and differentiable, we define C such that $\|f_0'\| < C < \infty$ and $\|f^{\mathrm{BS}}\| < C$. The measurable set of taut spline functions $L^*_{f^{\mathrm{BS}}} = \left\{ f^{\mathrm{BS}} : \|f_0 - f^{\mathrm{BS}}\|_\infty < \Delta C \right\}$ can be shown to exist (de Boor, 2001) and we define a map $\mathcal{G}: L^*_{f^{\mathrm{BS}}} \to L^*_{f^{\mathrm{LX}}}$ where $L^*_{f^{\mathrm{LX}}}$ a subset of all possible local extremum spline functions with H change points. Consider

$$\|f^{\mathrm{BS}} - f^{LX}\|_\infty = \sup_{x \in \mathcal{X}} |f^{\mathrm{BS}}(x) - f^{\mathrm{LX}}(x)| \quad \text{(A1)}$$

and let $\beta_0 = f^{\mathrm{BS}}(0)$. For the exactly H extrema $\alpha_1^{\mathrm{BS}} < ldots < \alpha_H^{\mathrm{BS}}$ in f^{BS} defined by the taut spline, set $\alpha_h = \alpha_h^{\mathrm{BS}}$. Additionally, if $f^{\mathrm{BS}}\left(\alpha_1^{\mathrm{BS}}\right) - f^{\mathrm{BS}}(0) \geq 0$ with H odd, then set $M = -1$; otherwise set $M = 1$. In the case where $f^{\mathrm{BS}}\left(\alpha_1^{\mathrm{BS}}\right) - f^{\mathrm{BS}}(0) < 0$ with H odd, then set $M = 1$ otherwise set $M = -1$.

Rewriting the right hand side of (A1) in a form based upon the derivative, we have

$$\sup_{x \in \mathcal{X}} \left| \int_{-\infty}^{x} \sum_{k=1}^{K+j-1} \kappa_k B_{(j,k)}(\xi) - \beta_k G(\xi) B_{(j,k)}(\xi)\, d\xi \right|,$$

$$\leq \sum_{k=1}^{K+j-1} \sup_{x \in \mathcal{X}} \left| \int_{\tau_k}^{x} \kappa_k B_{(j,k)}(\xi) - \beta_k G(\xi) B_{(j,k)}(\xi)\, d\xi \right|,$$

where the derivative of f^{BX} is based upon the derivative formula for B-Splines (de Boor, 2001) and $G(\xi) = \prod_{h=1}^{H} (\xi - \alpha_h)$.

Because of the taut spline construction of f^{BS}, we know that for all k, h such that $\alpha_h \notin [\tau_k, \tau_{k+j-1}]$ one has $\mathrm{sgn}(\kappa_k) = \mathrm{sgn}(G(x))$, for all $x \in [\tau_k, \tau_{k+j-1}]$. Here $\mathrm{sgn}(\cdot)$ is the signum function. On each of these intervals let

$$\beta_k = \frac{\int_{\tau_k}^{\tau_{k+j-1}} \kappa_k B_{(j,k)}(\xi)\, d\xi}{\int_{\tau_k}^{\tau_{k+j-1}} G(\xi)\, B_{(j,k)}(\xi)\, d\xi}.$$

As $B_{(j,k)}(x) \geq 0$, we have $\beta_k \geq 0$; further, one has

$$\int_{\tau_k}^{\tau_{k+j-1}} \kappa_k B_{(j,k)}(\xi) - \beta_k G(\xi) B_{(j,k)}(\xi)\, d\xi = 0$$

for all intervals such that $a_h \in [\tau_k, \tau_{k+j-1}]$.

For the at most H coefficients defined on splines that are nonzero in the intervals $a_h \in [\tau_k, \tau_{k+j-1}]$, set these coefficients to zero. As there are a finite number of intervals whose error is non-zero and f^{BS} is bounded, the maximum error is at most $(H+1)(j+1)\Delta C$ for any x and

$$\|f^{\mathrm{BS}} - f^{\mathrm{LX}}\|_\infty \leq (H+1)(j+1)\Delta C.$$

Consequently, for any ε, consider taut B-spline constructions on knot sets $\mathcal{T}$ such that $\Delta \leq \varepsilon[\{2(H+1)(j+1)\}C]^{-1}$ that also have $\|f_0 - f^{\mathrm{BS}}\|_\infty < \varepsilon/2$. Then one has

$$\|f_0 - f^{LX}\|_\infty \leq \|f_0 - f^{\mathrm{BS}}\|_\infty + \|f^{\mathrm{BS}} - f^{\mathrm{LX}}\|_\infty = \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

completing the proof. □

Proof of Lemma 1

The function $\mathcal{G}$ in Theorem 1 is measurable. If $L^*_{f^{\mathrm{BS}}}$ is measurable on some abstract measure space, one has $\mathrm{pr}\left(\|f^{\mathrm{LX}} - f_0\|_\infty < \varepsilon | \mathcal{T}_{\mathcal{M}}\right) > 0$ for any $\varepsilon > 0$ and some $\mathcal{T}_{\mathcal{M}}$. Given the prior puts probability over knot sets having knot spacings that are arbitrarily close, that is $\Delta \leq \varepsilon[\{2(H+1)(j+1)\}C]^{-1}$ as in Theorem 1, we conclude that

$\mathrm{pr}\left(\|f_0 - f^{\mathrm{LX}}\|_\infty < \varepsilon\right) = \mathrm{pr}\left(\|f^{\mathrm{LX}} - f_0\|_\infty < \varepsilon | \mathcal{T}_{\mathcal{M}}\right) \mathrm{pr}(\mathcal{T}_{\mathcal{M}}) > 0$ for all $\varepsilon > 0$. □

Proof of Theorem 2

We verify the conditions given in A1 and A2 of Theorem 1 of Choi & Schervish (2007). If there is positive prior probability, Lemma 1, within all neighborhoods of (f_0, σ^2), one can use Choi & Schervish (2007), section 4, to show that the conditions of A1 of Theorem 1 are met. To verify A2 we have that $\mathcal{F}^{H+}$ and $\mathcal{F}^{H+}$ are subsets of all continuous differentiable functions on $\mathcal{X}$ which were considered in Choi & Schervish (2007); consequently, we appeal to Theorem 2 and 3 of Choi & Schervish (2007) to construct suitable tests for both random and fixed designs using $W_{\varepsilon,n}$ and U_ε. We need only verify (iii) in part A2.

As in Choi & Schervish (2007), assume that $M_n = \mathcal{O}(n^\alpha)$ with $1/2 < \alpha < 1$. We show that $\mathrm{pr}\left(\|f^{\mathrm{LX}}(x)\|_\infty < M_n\right) \leq C_0 exp(-nC_1)$ and $\mathrm{pr}\left(\|f'^{\mathrm{LX}}(x)\|_\infty > M_n\right) \leq C_2 exp(-nC_3)$

for some C_0, C_1, C_2, $C_3 > 0$. Define $B^*_{(j,k,\mathscr{M},\alpha)}(X)$ as the design matrix given model $\mathscr{M}$ and a particular α figuration. Let $A = \sup_{\mathscr{M},k,\alpha,x} |B^*_{(j,k,\mathscr{M},\alpha)}(X)|$ and $K_{\mathscr{M}}$ be the number of spline coefficients in model $\mathscr{M}$ then

$$\begin{aligned}&\mathrm{pr}\left\{\|f^{LX}(x)\|_\infty > M_n\right\}\\ &= \int pr\left\{\|\sum_{k=1}^{K_{\mathscr{M}}}\beta_k B^*_{(j,k,\mathscr{M},\alpha)}(X)\|_\infty > M_n|\mathscr{M}\right\} d\alpha\, d\mathscr{M}\, d\pi\, d\lambda \le \int pr\left\{\sum_{k=1}^{K_{\mathscr{M}}}\|\beta_k B^*_{(j,k,\mathscr{M},\alpha)}(X)\|_\infty > M_n|\mathscr{M},\beta>0\right\} d\alpha\, d\mathscr{M}\, d\pi\end{aligned}$$

(A2)

Where the last inequality comes from the the Chernoff bounds.

Now let pr*($\mathscr{M}$) be the probability of a branching process where $\zeta < 0.5$ is constant for all children, then there exists a $\mathscr{K}$ such that $\{\mathrm{pr}^*(\mathscr{M})\}^2 \ge \mathrm{pr}(\mathscr{M})$ for all $\mathscr{M}$ such that $K_{\mathscr{M}} \ge \mathscr{K}$. Partition the sum into the finite sum where $K_{\mathscr{M}} < \mathscr{K}$ and the infinite sum $K_{\mathscr{M}} \ge \mathscr{K}$. As the finite sum is finite for all $0 < t < \lambda$, continuing with (A2):

$$\le exp(-M_n t)\int C_1 + \left[\sum^{K_{\mathscr{M}} \ge \mathscr{K}} \left(\frac{\lambda-\pi t}{\lambda-t}\right)^{K_{\mathscr{M}}} \{pr^*(\mathscr{M})\}^2\right] d\alpha\, d\pi\, d\lambda \le exp(-M_n t)\int C_1 + C_2\left[\sum^{K_{\mathscr{M}} \ge \mathscr{K}} \left(\frac{\lambda-\pi t}{\lambda-t}\zeta\right)^{K_{\mathscr{M}}} pr^*(\mathscr{M})\right] d\alpha\, d\pi\, d$$

where the last inequality exists as λ is bounded above zero, which implies one can choose some $t < \lambda$ such that $(\lambda - \pi t)/(\lambda - t)\zeta < 1$. This implies that

$$\mathrm{pr}\left\{\|f^{LX}(x)\|_\infty > M_n\right\} \le C_0 exp(-nC_1).$$

A derivation similar to the above can be used to show the same holds for $\mathrm{pr}(\|f'^{LX}(x)\|_\infty > M_n) \le C_2 \exp(-nC_3)$. One can find a $B = \sup_{\mathscr{M},k,\alpha,x} |B'^*_{(j,k,\mathscr{M},\alpha)}(X)|$ and substitute B for A and $B'^*_{(j,k,\mathscr{M},\alpha)}(X)$ for $B^*_{(j,k,\mathscr{M},\alpha)}(X)$ in the above derivation.

References

Baraud Y, Huet S, Laurent B. Testing convex hypotheses on the mean of a Gaussian vector. application to testing qualitative hypotheses on a regression function. Annals of Statistics. 2005:214–257.

Biller C. Adaptive Bayesian regression splines in semiparametric generalized linear models. Journal of Computational and Graphical Statistics. 2000; 9:122–140.

Bornkamp B, Ickstadt K. Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis. Biometrics. 2009; 65:198–205. [PubMed: 18510655]

Choi T, Schervish MJ. On posterior consistency in nonparametric regression problems. Journal of Multivariate Analysis. 2007; 98:1969–1987.

de Boor, C. A Practical Guide to Splines. New York: Springer Verlag; 2001.

DiMatteo I, Genovese CR, Kass RE. Bayesian curve-fitting with free-knot splines. Biometrika. 2001; 88:1055–1071.

Feller, W. Introduction to Probability Theory and Its Applications. Vol. I. New York: John Wiley and Sons; 1974.

Genz A. Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics. 1992; 1:141–149.

Genz A, Kwong KS. Numerical evaluation of singular multivariate normal distributions. Journal of Statistical Computation and Simulation. 2000; 68:1–21.

Geyer, CJ. Markov chain Monte Carlo maximum likelihood. In: Keramidas, EM., editor. Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Red Hook, NY: Interface Foundation of North America; 1991. p. 1-8.

Geyer, CJ. Importance sampling, simulated tempering and umbrella sampling. In: Brooks, S.Gelman, A.Jones, G., Meng, X., editors. Handbook of Markov Chain Monte Carlo. Boca Raton: Chapman & Hall/CRC; 2011. p. 295-311.

Godsill SJ. On the relationship between Markov chain Monte Carlo methods for model uncertainty. Journal of Computational and Graphical Statistics. 2001; 10:230–248.

Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Green, PJ., Silverman, BW. Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach. Boca Raton: CRC Press; 1993.

Hans C, Dunson D. Bayesian inferences on umbrella orderings. Biometrics. 2005; 61:1018–1026. [PubMed: 16401275]

Holmes C, Mallick B. Generalized nonlinear modeling with multivariate free-knot regression splines. Journal of the American Statistical Association. 2003; 98:352–368.

Lang S, Brezger A. Bayesian P-splines. Journal of Computational and Graphical Statistics. 2004; 13:183–212.

Lavine M, Mockus A. A nonparametric Bayes method for isotonic regression. Journal of Statistical Planning and Inference. 1995; 46:235–248.

Meyer M. Inference using shape-restricted regression splines. The Annals of Applied Statistics. 2008; 2:1013–1033.

Meyer MC, Hackstadt AJ, Hoeting JA. Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. Journal of Nonparametric Statistics. 2011; 23:867–884.

Neelon B, Dunson D. Bayesian isotonic regression and trend analysis. Biometrics. 2004; 60:398–406. [PubMed: 15180665]

Ramgopal P, Laud P, Smith A. Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve. Biometrika. 1993; 80:489–498.

Ramsay J. Monotone regression splines in action. Statistical Science. 1988; 3:425–441.

Salomond, JB. The Contribution of Young Researchers to Bayesian Statistics. New York: Springer; 2014. Adaptive Bayes test for monotonicity; p. 29-33.

Scott JG, Shively TS, Walker SG. Nonparametric Bayesian testing for monotonicity. Biometrika. 2015; 102:617–630.

Shi, JQ., Choi, T. Gaussian Process Regression Analysis for Functional Data. Boca Raton: CRC Press; 2011.

Shively T, Sager T, Walker S. A Bayesian approach to non-parametric monotone function estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009; 71:159–175.

Shively T, Walker S, Damien P. Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. Journal of Econometrics. 2011; 161:166–181.

Viboud C, Miller M, Olson DR, Osterholm M, Simonsen L. Preliminary estimates of mortality and years of life lost associated with the 2009 A/H1N1 pandemic in the US and comparison with past influenza seasons. Public Library of Science: Currents Influenza. 2010; 2:1–7.

Walker S, Damien P, Lenk P. On priors with a Kullback–Leibler property. Journal of the American Statistical Association. 2004; 99:404–408.

Wang JC, Meyer MC. Testing the monotonicity or convexity of a function using regression splines. Canadian Journal of Statistics. 2011; 39:89–107.

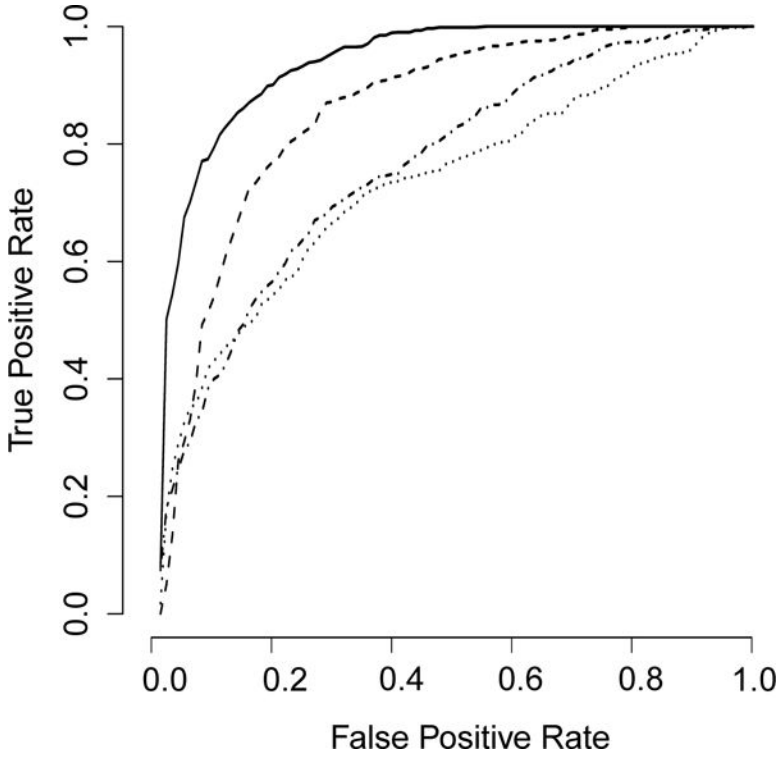

Fig. 1.

The receiver operating curve for the four tests defined for hypothesis $\mathbb{H}_1$ for all 1,400 simulations. The black line represents the local extremum spline, dashed line the approach of Salomond, dashed-dotted line the approach of Baraud, and dotted line the approach of Wang and Meyer.

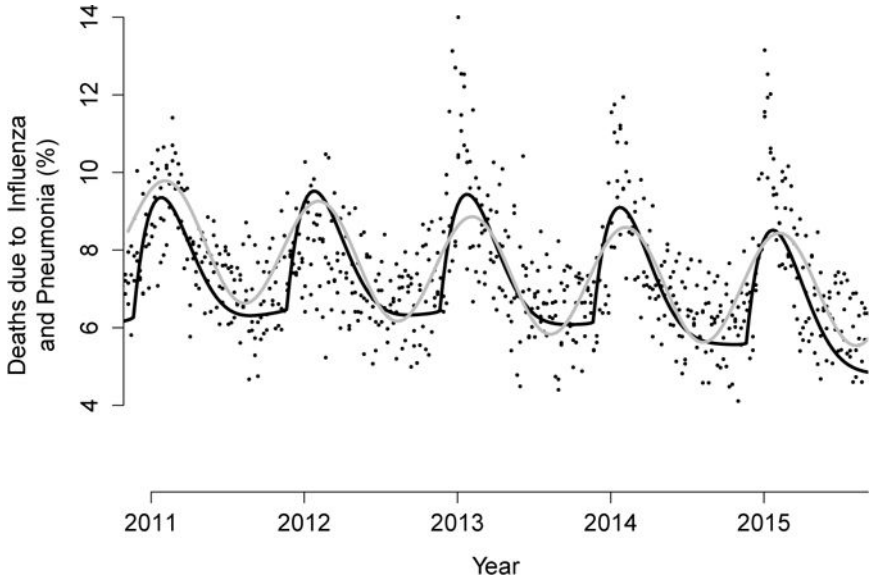

Fig. 2.

Estimate of the expected rate of seasonal influenza and pnuemonia deaths using the local extremum spline, black line, compared to the observed rate of influenza and pnuemonia deaths estimated using the Center for Disease Control's standard approach, gray line. Dots represent observed state level influenza and pneumonia percentages.

Table 1

Estimated mean squared error for all functions. For each function, the left value represents the simulation condition $\sigma^2 = 1$ and the right value represents the simulation condition $\sigma^2 = 4$. Asterisks signify that the number is significantly different than the local extremum spline at the one-sided 0·05 level.

True Function	Local Extremum Splines	Smoothing Splines	Bayesian P-Splines	Gaussian Process
f_1	1·60/0·49	2·11*/0·58	2·28*/0·55	2·15*/0·71*
f_2	2·59/0·09	4·19*/0·13*	3·82*/0·11*	5·26*/0·15*
f_3	1·57/0·49	2·43*/0·67*	2·26*/0·92*	2·64*/0·79*
f_4	1·70/0·49	2·10*/0·56*	2·15*/0·49	1·90*/ 0·59*
f_5	2·55/0·61	3·69*/1·12*	3·39*/0·98*	3·90*/1·14*
f_6	2·17/0·69	2·57/0·72	5·16*/0·72	2·44/0·79*
f_7	2·38/0·66	3·39*/1·05*	3·96*/0·85*	3·30*/0·90*

Table 2

Percent of samples where the model was correctly chosen as having two extrema, which is hypothesis $\mathbb{H}_3$, using a cut point of 6.

Function	***n***			
	100	200	300	400
g_7	78	90	98	96
g_8	14	32	22	46
g_9	76	88	98	100