

Statistical application of barycentric rational interpolants: an alternative to splines

Rose D. Baker · Dan Jackson

Received: 26 March 2013 / Accepted: 15 January 2014 / Published online: 4 February 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Spline curves, originally developed by numerical analysts for interpolation, are widely used in statistical work, mainly as regression splines and smoothing splines. Barycentric rational interpolants have recently been developed by numerical analysts, but have yet seen very few statistical applications. We give the necessary information to enable the reader to use barycentric rational interpolants, including a suggestion for a Bayesian prior distribution, and explore the possible statistical use of barycentric interpolants as an alternative to splines. We give the all the necessary formulae, compare the numerical accuracy to splines for some Monte-Carlo datasets, and apply both regression splines and barycentric interpolants to two real datasets. We also discuss the application of these interpolants to data smoothing, where smoothing splines would normally be used, and exemplify the use of smoothing interpolants with another real dataset. Our conclusion is that barycentric interpolants are as accurate as splines, and no more difficult to understand and program. They offer a viable alternative methodology.

Keywords Spline · Smoothing · Barycentric rational interpolation · Meta-analysis

1 Introduction

Regression splines are widely used in statistical analysis for curve-fitting, and have the advantage over nonparametric methods that likelihood functions can be written down,

R. D. Baker (✉)
School of Business, University of Salford, Salford, UK
e-mail: rose.baker@cantab.net

D. Jackson
MRC Biostatistics Unit, Cambridge, UK
e-mail: daniel.jackson@mrc-bsu.cam.ac.uk

parameters estimated and inference done for a variety of models, such as multiple linear regression, survival modelling, and logistic modelling (e.g. [Harrell 2001](#)). We are interested in evaluating a newly-developed alternative to spline modelling, the use of barycentric rational interpolants (BRIs). Hence we do not discuss further other techniques of nonparametric regression that aim to smooth curves, such as the loess method. Use of fractional polynomials ([Royston and Altman 1994](#)) is the only other technique that can be used for parametric inference about nonlinear relationships, but this although often useful does not have the flexibility of a spline curve. Hence we shall be presenting BRIs as an alternative to regression and smoothing splines.

It may seem odd that a method developed by numerical analysts for interpolating functions should be recommended for statistical use. The point is of course that one takes function values μ_k defined at regularly-spaced points x_k as unknown parameters to be estimated, so any methodology that can interpolate a function can also be used to fit a flexible functional form to statistical data.

Spline curves, originally developed by numerical analysts for function interpolation, have been used in statistics for many years and need little introduction. [Wegman and Wright \(1983\)](#) review the then state of the art, and useful books are [de Boor \(2001\)](#) and [Wahba \(1990\)](#). The use of splines is typically in parameterising dependence on a covariate, e.g. patient age in a medical study of the effect of some treatment intervention.

One needs a flexible curve that can represent an unknown functional form, where the function will often be ‘smooth’. A cubic spline represents the curve by a piecewise (continuous) cubic curve, with continuous first and second derivatives at the ‘knots’ where the value of the function is specified. Natural cubic splines impose the constraint that the second derivative is zero at the end knots.

One could say that splines although very useful have the drawback of computational complexity. There are several methods of fitting them. To fit a spline curve by least squares, one can introduce a new covariate for each knot, such that a conventional least-squares fit can be done using a statistics package. Computing these covariate values is feasible but requires the inversion of a tridiagonal matrix. A solution is the use of B (for basis) splines ([de Boor 2001](#)). The computations here are simpler, but now the regression coefficients are no longer the function values at the knots, but are ‘control points’. A third approach, use of the truncated power basis, is simple to program but is numerically unstable when there are many knots, and still does not directly yield estimates of the covariate value at the knots.

There is now a newer method of function interpolation, barycentric rational interpolation. Its performance for function interpolation compares favourably with splines ([Floater and Hormann 2007](#)). Although over the last few years the properties of BRIs as interpolants of smooth functions have been studied extensively by numerical analysts, barycentric interpolation has been hardly at all used in statistical work. This paper seeks to introduce BRIs to the statistical community, and to explore whether a BRI fit can be a competitor to a spline fit, and how it might be used. Our main methodology is least-squares fitting to Monte-Carlo data, and we also fit two real datasets. This compares BRIs with regression splines; subsequently, with the aid of a third example we also explore the use of smoothing and penalized BRIs, where there are many fitted parameters plus a roughness penalty.

2 Barycentric interpolation

The barycentric form for an interpolant $f(x)$ between coordinates (x_k, μ_k) where $x_1 < x_2 < \dots < x_m$ has been known at least since [Taylor \(1945\)](#), who showed that a polynomial interpolant could be written in barycentric form. The formula with m nodes is

$$f(x) = \frac{\sum_{k=1}^m w_k \mu_k / (x - x_k)}{\sum_{k=1}^m w_k / (x - x_k)}, \tag{1}$$

where we differ from some texts in summing from unity rather than from zero. The weights w_k are arbitrary and the curve will pass through the points (x_k, μ_k) for any vector of weights \mathbf{w} . The origin of the odd name ‘barycentric’ (centre of mass) now becomes clearer: the points μ_k are weighted, as in the formula for the centre of mass or for a sample mean; barycentric coordinates were introduced by Möbius in 1827. Evaluation of $f(x)$ is easy, and it is only necessary to check whether $|x - x_i| < \epsilon$ for any i , with (typically) $\epsilon = 10^{-8}$. If so, $f(x) \simeq \mu_i$, and otherwise one can safely compute from (1).

BRIs have so far scarcely been used in statistics. [Baker and McHale \(2012\)](#) used BRIs to study the long-term performance of golfers, and [Jackson et al. \(2013\)](#) used BRIs in meta-analysis.

2.1 Another method of computation

We introduce another method of computation of $f(x)$, that is also useful for computing derivatives. With x_i the closest node to x , i.e. $|x - x_i| < |x - x_k| \forall_{k \neq i}$ multiply numerator and denominator of (1) by $x - x_i$ and write $x - x_i = x - x_k + (x_k - x_i)$ to obtain

$$f(x) = \frac{\sum_{k=1}^m w_k \mu_k + \sum_{k \neq i} (x_k - x_i) w_k \mu_k / (x - x_k)}{\sum_{k=1}^m w_k + \sum_{k \neq i} (x_k - x_i) w_k / (x - x_k)}, \tag{2}$$

which now poses no computational problem when $x \simeq x_i$.

Derivatives can be computed, and differentiating (2) and setting $x = x_i$ yields the formula for the first derivative

$$df/dx|_{x=x_i} = -(1/w_i) \sum_{j \neq i} \frac{w_j (\mu_i - \mu_j)}{x_i - x_j}, \tag{3}$$

which we shall use later in a statistical test. [Klein and Berrut \(2012\)](#) give a recursive formula for calculating higher derivatives.

2.2 Computing the weights

Polynomial interpolants such as the Lagrangian interpolant can be put into barycentric form (e.g. [Berrut 1988](#)). However, the fact that any set of weights \mathbf{w} will give an interpolated curve that passes through the specified points opens the way for other interpolants besides polynomials. [Berrut \(1988\)](#) gave weights for a barycentric rational

interpolant, where a ratio of polynomials is used to interpolate. Berrut's interpolant was free of poles, a potential problem with ratios of polynomials. This approach was developed further by Floater and Hormann (2007) who derived sets of weights giving accuracies of $O(h^{d+1})$ where d is the order of the interpolant and h the largest step size (largest gap between consecutive values of x).

The formula for constructing weights is

$$w_k = (-1)^k \sum_{i=k-d}^k \prod_{j=i, j \neq k}^{i+d} \frac{1}{|x_k - x_j|}, \quad (4)$$

where an empty product is taken as unity, and Press et al. (2007) give C++ code. Using this formula, the weights of order zero are $w_k = (-1)^k$, and weights of order one are

$$w_k = (-1)^k \left\{ \frac{1}{x_k - x_{k-1}} + \frac{1}{x_{k+1} - x_k} \right\},$$

with terms with out-of-range values x_0 or x_{n+1} being omitted. For equally-spaced values of μ_k these weights with $d = 1$ reduce to $w_k = (-1)^k$, where now the end nodes have weight halved. This recalls the halving of end-point weights in the trapezoidal rule for integration formula. We normally use equally-spaced nodes in this paper, but keep the general form in all formulae.

2.3 Interpreting BRIs

The d th order interpolant can reproduce polynomials of degree at least d . It turns out that the $d = m - 2$ interpolant is a polynomial of degree $m - 1$, so there is no point taking $d > m - 2$.

It is helpful in understanding these interpolants to examine the first few. With $m = 1$, (1) reduces to $f = \mu_1$, a constant, and for $m = 2$ to a straight line. The $d = 0$ and $d = 1$ models are the same, with weights 1, -1 . For $m = 3$ there are two models. The $d = 1$ model is a quadratic, with weights $1/2, -1, 1/2$ for equally-spaced values of x , and the $d = 0$ model is the first interpolant that is a ratio of polynomials, being a ratio of quadratics. The value of f tends to a constant at large $|x|$, in fact to $\mu_1 - \mu_2 + \mu_3$. It can be verified from (3) that the slopes are equal and opposite at x_1 and x_3 under equal-spacing. The $m = 4$ interpolant with $d = 0$ is linear in the tails, with the same slope in both tails. The $d = 1$ interpolant with weights $(1/2, -1, 1, -1/2)$ is also linear in the tails, and the $d = 2$ interpolant with weights $(1/3, -1, 1, -1/3)$ is a (cubic) polynomial.

2.4 Periodic or circular data

There is also a barycentric interpolant for periodic or circular data, such as wind-speed (e.g. Berrut and Trefethen 2004; Berrut and Welscher 2007). It is

$$f(\phi) = \frac{\sum_{k=1}^m (-1)^k \operatorname{cst} \frac{\phi - \phi_k}{2} \mu_k}{\sum_{k=1}^m (-1)^k \operatorname{cst} \frac{\phi - \phi_k}{2}}$$

where

$$\operatorname{cst}(\phi) = \begin{cases} \cot \phi & \text{if } m \text{ even} \\ \operatorname{cosec} \phi & \text{if } m \text{ odd.} \end{cases}$$

2.5 Extension to multiple covariates

When there are several covariates, the BRIs generalise in the same way as do splines, i.e. for the bivariate case, writing weights for x as w_k and for z as v_l ,

$$f(x, z) = \frac{\sum_{k=1}^{m_x} \sum_{l=1}^{m_z} w_k v_l \mu_{kl} / (x - x_k)(z - z_l)}{\sum_{k=1}^{m_x} w_k / (x - x_k) \sum_{l=1}^{m_z} v_l / (z - z_l)}.$$

This introduction has given the essentials needed for use of barycentric interpolation. Before fitting data, we try to gain some insights into the nature of barycentric interpolants that bear on their statistical use.

3 A comparison of BRIs with splines

Unlike spline curves, barycentric curves are infinitely differentiable, which is a nice property. This observation suggests that BRIs might be slightly better than splines for fitting smooth functions, but the functions encountered in statistical work may not be smooth, and may have a step or ramp. For example, incidence of deaths among young people from motor accidents might be expected to increase sharply above the legal age for driving. One would model such obvious discontinuities explicitly, but others might go unremarked. We shall see how well BRIs cope with discontinuities later. It is of course not hard to model a step or a ramp using BRIs if wished. For a step, one uses two interpolants (1), one either side of a common node, and for a ramp one makes the values of μ at the common node equal.

Floater and Hormann (2007) used the construction of blended polynomials of order d to derive optimal weights and to show the absence of poles. They took polynomials $p_i(x)$ of degree $\leq d$ that interpolate f at the $d + 1$ points $x_{i+1} \cdots x_{i+d+1}$. Then they let

$$f(x) = \frac{\sum_{i=0}^{m-d+1} \lambda_i(x) p_i(x)}{\sum_{i=0}^{m-d+1} \lambda_i(x)}$$

where

$$\lambda_i(x) = \frac{(-1)^i}{(x - x_{i+1}) \cdots (x - x_{i+d+1})},$$

and they showed that this expression can be rewritten in barycentric form. Now each polynomial $p_i(x)$ interpolates only $d + 1$ adjacent values of f .

Their derivation shows that barycentric interpolation has a kinship with spline interpolation; local polynomials are used in both methods, albeit differently. In spline fitting, the local polynomial alone determines $f(x)$ within a range (x_i, x_{i+1}) , and it must be made to join as seamlessly as possible to the next segment at the end of its range. In barycentric interpolation, the local polynomial contributes to μ at only $d + 1$ adjacent values of x . The polynomial's influence on $f(x)$ continues however outside that range, because λ_i does not have local support, giving infinite differentiability. For full details, see Floater and Hormann (2007).

Since their introduction, the properties of the interpolants have been studied in detail. It has been rigorously shown as one might anticipate that the k th derivative has error $O(h^{d+1-k})$ for stepsize h , order of interpolation d (Klein and Berrut 2012), and that the integral has error $O(h^{d+2})$ (Klein and Berrut 2010), Bos et al. (2012). Note that the natural cubic spline has error $O(h^2)$.

Conceptually, BRIs seem mysterious in a way that splines are not; cubic spline curves can be imagined as a (simplified) mathematical model of the wooden splines formerly used by draftsmen, the cubic polynomial used between knots minimising an approximation to the energy of curvature of a wooden spline. In fact, the cubic minimises $\int y''(x)^2 dx$ on applying the Euler-Lagrange equation, whereas the energy of curvature is a more complex function. Currently there is no corresponding variational derivation of BRIs, and such a physical interpretation would be helpful.

4 Choosing an appropriate value of d

We might naively think that the order d should be chosen as large as possible, if increasing d increases interpolation accuracy. This cannot be correct, because when $d = m - 2$, we regain the polynomial interpolant, and polynomials are known to perform badly for equally-spaced points, because of the Runge phenomenon. With large d , the interpolant is a blending of high-order polynomials, and so can oscillate unduly between nodes, especially near the ends of the range.

Figure 1 shows barycentric interpolant fits with d from 0 to 3, and a spline fit to the points (0,1), (0.25,5), (0.5,0), (0.75,2) and (1,4).

It can be seen how the $d = 0$ fit is essentially linear between the nodes, although smooth, and how increasing d increases the size of the oscillations between nodes. The $d = 3$ line is the quartic polynomial interpolant. The spline fit is usually close to the $d = 1$ curve, although at the right hand side, it increases linearly, whereas all the barycentric interpolant curves decrease. This is because the discontinuous third derivative of the spline curve allows it to 'take a short-cut' in a way that the infinitely differentiable BRI curves do not.

Güttel and Klein (2012) show how rounding error and random error lower the optimal value of d for large m , using the concept of the Lebesgue constant. The optimal choice of d in statistical work clearly depends on the nature of the function being fitted; if for example it is a polynomial, we need $d = m - 2$.

The obvious frequentist way to choose d is to find the value of d that maximises the log-likelihood of the fit, and to condition the inference on that value of d . Increasing

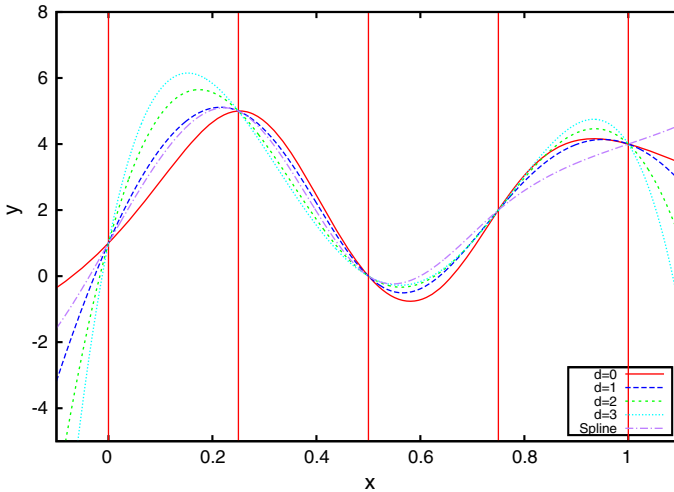


Fig. 1 Rational barycentric interpolant and natural spline curves with 5 nodes/knots (vertical lines)

the value of d does not of course introduce any fresh parameters. For a related problem and solution with splines, see [Kauermann and Opsomer \(2011\)](#).

A Bayesian approach would be to put a prior distribution on d . This makes sense, because to choose a value of d is to make a (subtle) statement about the functional form of the dependence of the predictand on the covariate. The curves are always smooth, but as d increases, what computer-aided designers call the ‘fairness’ of the curve decreases, and its length increases, i.e. the curve becomes more wiggly. This is what statisticians working in curve-fitting call ‘roughness’, so the curve becomes ‘rougher’. This is a different concept from mathematical roughness. Barycentric interpolation always produces curves that are not ‘rough’ in the mathematical sense, being infinitely differentiable; see e.g. [Sapidis and Farin 1990](#).

The minimum value of d is 0, and the maximum is $m - 2$, which already gives a polynomial interpolation, because choosing $m - 1$ gives the same weights. This follows from theorem 2 of [Floater and Hormann \(2007\)](#) as pointed out by [Klein \(2011\)](#).

We suggest a suitable prior would be to take d as following a binomial distribution with probability $p = 0.2$, sample size $m - 2$, so

$$\Pr(d) = \binom{m - 2}{d} p^d (1 - p)^{m - 2 - d}.$$

This is based both on the function interpolations of [Güttel and Klein \(2012\)](#) and on our own explorations. The optimal value of d is roughly proportional to m , with $m - 2$ as the maximum, and the nearest integer to $0.2(m - 2)$ was always a reasonable choice of d for the datasets we examined. Doubtless this prior could be improved for specific areas as further experience is gained.

It should be noted that, as can be seen from [Fig. 2](#), like spline fits, (1) is poor for extrapolating beyond the fitted range.

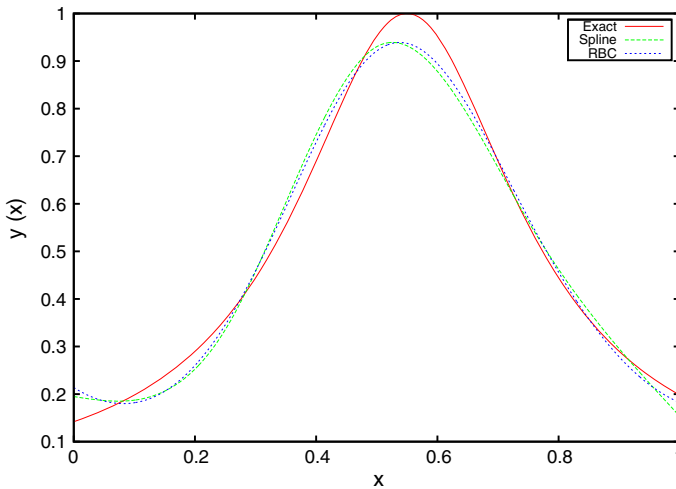


Fig. 2 The function $\{1 + 20(x - 0.55)^2\}^{-1}$ with spline fit from sample size 1,000 and the corresponding BRI fit (5 knots/nodes)

At large $|x|$ (1) reduces to a polynomial of order d (or $d + 1$ if $m - d$ is even). One would expect fits with higher d to extrapolate worse, and the relatively low weights given to the end points do not augur well for forecasting, where one wishes to discount earlier rather than later observations. A simple short-term forecast could be obtained by extrapolating linear trend using (3). The only advantage over a spline fit as usually carried out is that the values μ_i are immediately available from the least-squares fit, complete with a covariance matrix, so that forecasts can readily be made using for example Holt's method.

5 Comparison of BRIs and splines using Monte-Carlo datasets

We compare natural spline fits, which have error $O(h^2)$ for step-length h with BRI fits taking $d = 1$, which also have error $O(h^2)$, where h is stepsize. Table 1 shows mean absolute error (MAE) and root mean squared error (RMSE) for spline and BRI least-squares fits to ten functions $f(x)$, with data generated using the regression model $y_i = f(x_i) + \epsilon_i$, $\epsilon_i \sim N[0, \sigma^2]$.

We chose $\sigma = 0.1$. The 1,000 simulated x values were generated randomly in $(0, 1)$. The spline fits in Table 1 had $m = 5$ knots and the BRI fits 5 nodes, the knot/node points being equally spaced between 0 and 1. The rationale here is to mimic use of 'default' fitting options; 5 knots will be enough for most applications, and the setting $d = 1$ was used for the barycentric interpolants, as we have found this to be a good 'default' setting, and also because natural cubic splines and $d = 1$ BRI fits have the same order of error.

The procedure followed for both splines and BRIs was to generate variables $x_{i1} \cdots x_{im}$ such that the regression model is approximated as $y_i = \sum_{j=1}^m x_{ij} \mu_j + \epsilon_i$. For BRIs, with $f(x_i)$ substituted in from equation (1) we have

Table 1 Mean average errors (MAE), root mean square errors (RMSE) and Akaike information criterion (AIC) for spline and BRI fits with $d = 1$ to ten functions where $0 \leq x \leq 1$. using 5 knots/nodes. The function $H(x)$ denotes the Heaviside step function (zero if $x < 0$, else unity), and $(x - 1/2)_+$ is a ramp, i.e. $x - 1/2$ for $x > 1/2$, else zero. Sample size is 1,000

Function	MAE (sp)	MAE (BRIs)	RMSE (sp)	RMSE (BRIs)	AIC (sp)	AIC (BRIs)
$\frac{1}{1+20(x-0.55)^2}$.03000	.02477	.03614	.02997	-1602.67	-1637.37
$\sin(\pi x/2)$.00988	.00985	.01130	.01124	-1722.70	-1723.03
$\exp(-x)$.00961	.00939	.01120	.01104	-1723.00	-1722.77
$\exp(-2x^2)$.00959	.00964	.01157	.01255	-1721.65	-1716.54
$\ln(1 + 2x)$.00950	.00902	.01136	.01127	-1722.16	-1721.10
$\exp(-2(x - 1/2)^2)$.00954	.00927	.01119	.01116	-1722.83	-1722.35
$x \exp(-x)$.00940	.00900	.01124	.01110	-1722.59	-1721.85
$x + x^2 + x^3$.00945	.00919	.01526	.01222	-1720.72	-1718.02
$H(x - 1/2)$.12904	.11127	.18062	.16425	-278.12	-414.72
$(x - 1/2)_+$.01031	.01006	.01238	.01242	-1715.15	-1714.17

$$y(x_i) = \frac{\sum_{j=1}^m w_j \mu_j / (x_i - x_j)}{\sum_{k=1}^m w_k / (x_i - x_k)} + \epsilon_i = \sum_{j=1}^m x_{ij} \mu_j + \epsilon_i,$$

where now x_i is the i th sample value of x , and where

$$x_{ij} = \frac{w_j / (x_i - x_j)}{\sum_{k=1}^m w_k / (x_i - x_k)}. \tag{5}$$

The weights w_j were found using the formula of Floater and Hormann (2007), our Eq. 4.

Use of this procedure means that after creating these variables, a conventional linear regression can be performed, and this is more computationally efficient than evaluating \hat{y}_i using (1) each time it is needed.

For the spline model, natural cubic splines were generated for each x_i following the procedure described in Press et al. (2007), p 120–122. They describe the solution of a system of $m - 2$ linear equations, where the matrix is tridiagonal. As we are solving for the x_{ij} and the μ_j are unknown, it was necessary to invert the matrix. This was done in time $O(m^2)$ using the formula for the inverse of a tridiagonal matrix given by Usmani (1994).

The ten functions were simply what came to mind, adjusted to show a large variation in the range (0, 1). The first function, essentially the Cauchy pdf, is also the function famously used by Runge to show the oscillation of polynomial interpolants with increasing m . It can be seen that for nine of the ten functions, the MAE was slightly smaller using BRIs than using splines, and the RMSE was slightly smaller for two functions. The difference is small, and we conclude that BRIs perform at least as well as splines or slightly better in most cases; one could use either technique. This is consistent with Floater and Hormann (2007), who found similar results when interpolating in tables.

Table 2 Mean average errors (MAE) and root mean square errors (RMSE) for spline and BRI fits with $d = 1$ to the function $\exp(-1.5x) \cos(15x)$ using 10 knots/nodes. Optimal results are starred. Sample size is 1,000

Method	MAE	RMSE	AIC
Spline	.02039	.02686	-1651.49
BRI, $d = 0$.02215	.02903	-1640.51
BRI, $d = 1$.01774	.02228	-1678.12
BRI, $d = 2$.01924	.02360	-1683.31
BRI, $d = 3$.01314	.01598	-1711.22*
BRI, $d = 4$.01169*	.01467	-1710.33
BRI, $d = 5$.01218	.01480	-1706.38
BRI, $d = 6$.01212	.01371*	-1708.26
BRI, $d = 7$.01248	.01535	-1704.68
BRI, $d = 8$.01449	.01815	-1697.41

Interestingly, although both splines and BRIs perform poorly with the two non-smooth functions, the step function and the ramp, again the BRIs perform slightly better. The greater smoothness of the BRI (infinitely differentiable) does not count against it.

Figure 2 shows a plot of the first function in Table 1, $y = \{1 + 20(x - 0.55)^2\}^{-1} + \epsilon$, together with curves computed from the spline fit and from the BRI fit. It can be seen that both fits look equally good (or bad).

We also studied the effect of knot/node placement, by choosing the 3 central nodes randomly in $(0, 1)$. The error of both methods in general increased, but they performed similarly.

To explore the choice of optimum values of d for the BRIs, we fitted $y = \exp(-1.5x) \cos(15x) + \epsilon$ using $m = 10$ equally-spaced knots/nodes. To be fair to the spline method, it is of course possible to choose higher and higher order splines in the same way to obtain greater accuracy. However, this is not normally done, so we simply compare the BRI results with cubic spline fits. Table 2 shows the results.

The AIC is a widely-used measure of model adequacy that penalizes models with many parameters (Akaike 1974); lower AIC denotes a better fit.

The minimum-Akaike information criterion (MAICE) fit shows optimum $d = 3$. In terms of error, this is in reality dominated by fits with d from 4 to 7, but the improvement in error is small on moving to these higher values of d . The $d = 0$ fit is slightly worse than the spline fit, and for $d \geq 1$ the BRI performs better. The data presented are typical of the more extensive range of studies we carried out.

5.1 The ‘extended Floater–Hormann’ family of interpolants

Klein (2011) has produced the ‘extended Floater–Hormann’ family of interpolants. Briefly, using equally-spaced steps, one extrapolates the curve d steps beyond each end, using a Taylor series expansion, with derivatives computed from the BRI curve, e.g. as in (3). One then interpolates in the extended set of $m + 2d$ nodes using the (new) weights appropriate to an interpolation with $m + 2d$ nodes. This procedure reduces the Lebesgue error. A study of this was made using the ten datasets in Table 1 and

also with $d = 1$. The MAE and RMSE sometimes increased slightly and sometimes decreased slightly, with the spline fit now slightly better for the MAE of one function, and for the RMSE of another. The fits to the Heaviside (step) function and to the ramp deteriorated. This topic would bear further research, but it seems that the extra complexity of this method, and its poorer performance on discontinuous functions (for which it was not designed) indicate that it should not yet be used in statistical work.

6 Analysis of two real datasets

The aim here is to illustrate the use of the interpolants, and to show that the fit to data is as good as a spline fit would be. For both datasets the covariate is time in years, but the outcome variable differs.

6.1 Grand Slam tennis

The first example was drawn from the Grand Slam Tennis Archive, <http://www.tennis.ukf.net/>. Jimmy Connors played 276 grand-slam tennis matches over his professional career (1970–1992), and the probability of his winning a game can be estimated from the scores in each match. BRIs can be used to see how this probability changed over time.

The logit $y = \ln \hat{p}/(1 - \hat{p})$ was found for each match and the estimated standard error σ of the logit, where $\sigma^2 = \frac{1}{n\hat{p}(1-\hat{p})}$, n being the number of games played. These data were fitted to a BRI model, where both the number of nodes m and the parameter d were chosen to minimise the AIC. The log-likelihood function was

$$\ell(\boldsymbol{\mu}) = -(1/2) \sum_{i=1}^N (y(t_i) - \hat{y}(t_i))^2 / \sigma_i^2$$

to a constant, where N is the number of matches, $\hat{y}(t_i)$ the BRI prediction at time t_i , and σ_i the (known) corresponding value of σ . Nodes were spaced equally over the interval from first to last matches. The MAICE model was $m = 8, d = 3$, although the $d = 2$ model had an almost identical AIC. Figure 3 shows the result, where the logits and the fit have been back-transformed to probabilities of winning a game so as to be more easily interpretable. The 95 % CI is also shown.

The covariance matrix on the fitted model parameters is as usual the inverse of the Hessian matrix, $(\mathbf{V}^{-1})_{ij} = -\partial^2 \ell / \partial \mu_i \partial \mu_j$. As this is a linear regression, the standard error $\sigma(t)$ of $y(t)$ can be found from the covariance matrix \mathbf{V} on the fitted parameters, and the interval taken as $\hat{y}(t) \pm 1.96\sigma(t)$, and also back-transformed.

It can be seen that there are peaks around 1975 and 1983, and a smaller peak at 1990, which correspond to the high points of Connors’ career as seen for example from the Wikipedia entry. The spline fit (not shown) with the same knot positions had an AIC smaller by 0.13, so both fits are equally good. The fit to the model is not however too good, with $\chi^2[268] = 467.29$, where the goodness of fit chi-squared is computed from the predicted and expected probabilities of winning a game for each match. The

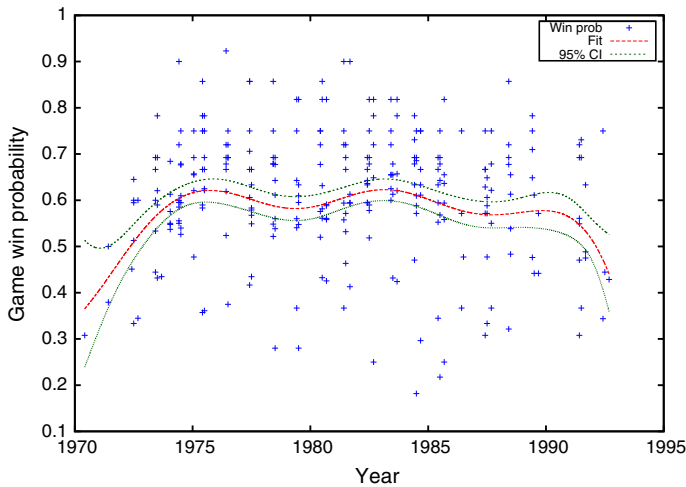


Fig. 3 Estimated game win probabilities for Jimmy Connors from 1970 to 1992, with the MAICE BRI fit and 95 % CI. Standard errors not shown for clarity

poor fit may be because the calibre of opponent was changing from game to game. Baker and McHale (2012) fitted BRI models to data on golf tournaments where the strength of each golfer was modelled using BRIs to avoid this problem. In the next section, we carry out a meta-analysis, where we will use a random effects model to address this problem.

6.2 Narcissism in college students

Twenge et al. (2008) carried out a meta-analysis of 85 samples of American college students who completed the Narcissistic Personality Inventory between 1979 and 2006. They found that narcissism levels were increasing in recent years. Their paper gives the mean scores, standard deviations and sample sizes, from which on removing studies with no quoted standard deviations, $n = 76$ studies remained. We carried out a random effects meta-analysis, using BRIs and choosing the number of equally-spaced nodes and the value of d to minimise the AIC. The log-likelihood was

$$\ell(\boldsymbol{\mu}, \tau) = (-1/2) \sum_{i=1}^n \{(y(t_i) - \hat{y}(t_i))^2 / (\sigma_i^2 + \tau^2) - (1/2) \ln 2\pi(\sigma_i^2 + \tau^2)\},$$

where τ^2 is the between-study or heterogeneity variance as used in random-effects meta-analysis.

The chosen model had 5 nodes and $d = 1$, and the AIC was 0.05 higher than for the best spline fit. The random effect error size was $\tau = 0.455 \pm 0.133$, giving an $I^2 = 81\%$ (Higgins and Thompson 2002). Figure 4 shows the data with error bars staggered within their year for clarity, the fitted line and the 95 % CI. The BRI analysis reveals the highly nonlinear temporal variation of the narcissism score, which initially dips slightly before rising steadily, and more recently rising at a faster rate. The initial

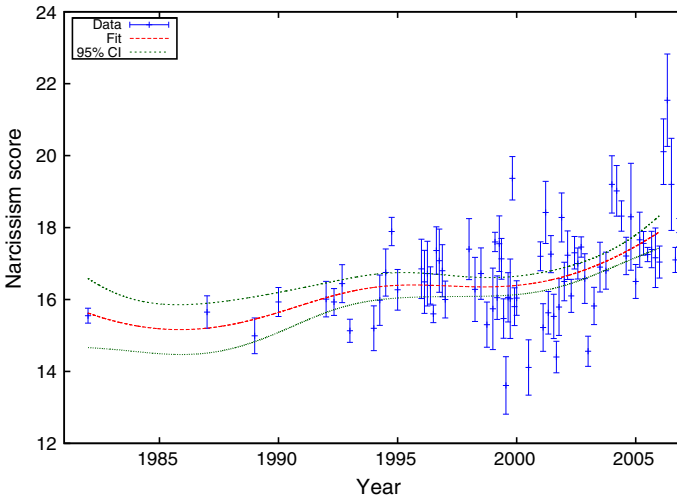


Fig. 4 Narcissism scores with standard errors from 1980 to 2005, with the MAICE BRI fit and 95 % CI. Errorbars staggered for clarity

dip arises solely because of one early influential observation and may be spurious, but the essential point is that BRIs can reproduce the varying slopes of the data adequately.

One can test whether the exit slope is zero by a Wald test for the slope (3) at the latest node, i.e. the date of the most recent study, using its computed standard error. Rewriting (3) as $df/dx|_{x=x_i} = \sum_{k=1}^m c_k \mu_k$, the standard error σ is $\{\sum_{i=1}^m \sum_{j=1}^m V_{ij} c_i c_j\}^{1/2}$. For an ordinary least squares regression, the standardised slope would follow a t-distribution with $n - m$ degrees of freedom, but as τ was also fitted, we take the standardised slope as asymptotically normal, as in conventional meta-analysis. The matrix \mathbf{V} is computed from the Hessian for all model parameters, i.e. $\boldsymbol{\mu}$ and τ .

The slope test is here hardly necessary, but the slope was 0.4318, standard error 0.1042, giving $Z = 4.14$, showing a significantly positive slope. Narcissism was increasing in 2006 at the rate of over 0.4 points per year. The corresponding initial slope was -0.235 ± 0.159 , giving $Z = -1.48$; narcissism is initially decreasing, but this is not statistically significant. Note that a fit with $d \geq 1$ should preferably be used, because the $d = 0$ interpolants can not reproduce a straight line.

This simple test could be useful in testing whether results from a meta-analysis are stable, an issue that has attracted some interest (e.g. Mullerleile and Mullen 2006). This type of analysis removes an objection to the analysis of Baker and Jackson (2010) which forced temporal trends to be essentially linear.

From these examples, we conclude that BRIs fit data as well as regression splines do, and give very similar results. We now turn to smoothing BRIs.

7 Smoothing BRIs

With a smoothing or penalised spline fit, ‘too many’ knots μ_i are used, and the function S minimised, where

$$S = \sum_{i=1}^n (y_i - \hat{y}(x_i))^2 + \lambda \int_{x_1}^{x_n} \hat{y}''(x)^2 dx, \tag{6}$$

λ is a constant controlling the size of the ‘roughness’ penalty, and the range of data is (x_1, x_n) . With a knot at every data point, we have a smoothing spline, and with fewer knots, a penalized spline regression.

When $\lambda = 0$ we regain an ordinary least squares spline fit for the $\hat{\mu}_i$. The integral is larger for ‘rough’ or wiggly fits, and the value of λ is chosen either manually, or if automatically either by using cross-validation to minimise the prediction error, or by choosing a minimum modified AIC solution (Harrell 2001). There are variants of this scheme; Eilers and Marx (1996) use a B-spline basis, equally-spaced knots and a difference rather than an integral roughness penalty, while Ruppert et al. (2003) use truncated power functions, knots based on quantiles, and an integral penalty as above.

If using an integral penalty as in (6), $y(x)$ and $y''(x)$ are linear in the μ_i , the integrand $y''(x)^2$ is a quadratic form in μ and so the integral can also be written in the quadratic form $\mu^T \mathbf{M} \mu$, where the roughness penalty matrix \mathbf{M} must be determined numerically. Then writing $\hat{y}(x_i) = \sum_{j=1}^m x_{ij} \mu_j$ we have that

$$\hat{\mu} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{M})^{-1} \mathbf{X}^T \mathbf{y}, \tag{7}$$

where $(\mathbf{X})_{ij} = x_{ij}$, showing the shrinkage of the estimators $\hat{\mu}$ resulting from the roughness penalty.

There is a great deal written about smoothing and penalized splines, but we merely wish to point out that a BRI, like a spline, is linear in the unknown μ_i , so the whole derivation of (7) goes through in an analogous way, with spline knots replaced by BRI nodes, and with x_{ij} as defined in (5).

To make the methodology work, it is necessary to describe the computation of the matrix \mathbf{M} for BRI fits in some detail. The matrix can be computed as follows: to avoid numerical problems, use the form (2) for the interpolant $\hat{y}(x)$, which is now (1) multiplied through top and bottom by $(x - x_s)$, where x_s is the nearest node to x . Then writing $N(x)$ for the numerator of (2) and $D(x)$ for its denominator, on differentiating $\hat{y}(x) = N(x)/D(x)$ we have $\hat{y}'' = N''/D - 2N'D'/D^2 - ND''/D^2 + 2ND'^2/D^3$. Writing \mathbf{M} as the outer product

$$M_{ij} = \int_{x_1}^{x_n} w_i w_j b_i(x) b_j(x) dx, \tag{8}$$

where $w_i b_i(x)$ is the coefficient of μ_i in $\hat{y}''(x)$, we have that

$$\begin{aligned} b_i(x) = & (x_i - x_s)[2D(x)^{-1}(x - x_i)^{-3} + 2D(x)'D(x)^{-2}(x - x_i)^{-2} \\ & + \{2(D(x)')^2 D(x)^{-3} - D(x)''D(x)^{-2}\}(x - x_i)^{-1}] \\ & + \{2(D(x)')^2 D(x)^{-3} - D(x)''D(x)^{-2}\}, \end{aligned} \tag{9}$$

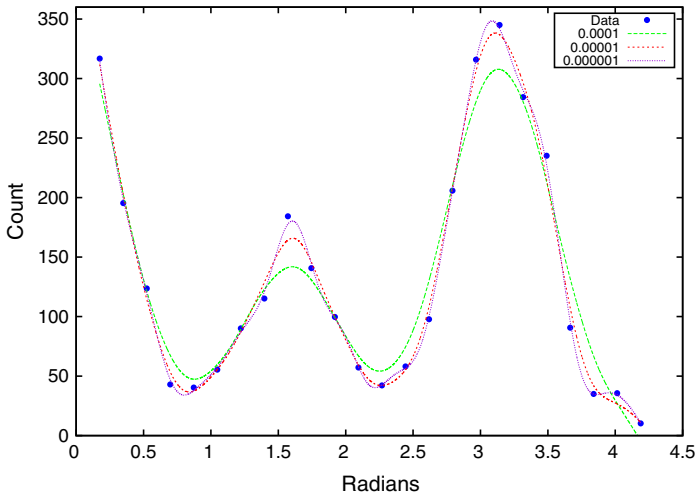


Fig. 5 Smoothing BRI fit to the C_{12} α -emission data at 21.6 MeV of DeAngelis et al. (1995). The data and fitted curves for the values of λ shown in the key are displayed

where the term in square brackets is not computed if $i = s$. Given these $b_i(x)$ the matrix \mathbf{M} is found from (8) by numerical integration. It is most efficient to nest the integration loop inside the loops that run over the suffices i, j . The integrand is smooth, so a simple method such as the trapezoidal rule or Simpson’s rule with say 1,000 nodes can be used. Otherwise there are more efficient specialised integration routines for integrating a vector of similar functions, e.g. the Numerical Algorithms Group (NAG) routine D01EAF. The $b_i(x)$ all oscillate differently with x , so if using adaptive integration, each element of \mathbf{M} would require different treatment, and it seems best to keep to a simple integrator with regularly-spaced nodes. The numerical values of the matrix \mathbf{M} is a function only of the number of nodes and their placement. With this formulation of the smoothing problem, the integrations need be done only once, before solving for the $\hat{\mu}_i$, which may be an iterative procedure if (6) is varied from being an ordinary least squares fit.

The matrix \mathbf{M} is symmetric, so only $m(m - 1)/2$ elements need be computed. Row and column sums are always zero, as are the linearly weighted sums $s_i = \sum_{j=1}^m M_{ij}x_j$, because curvature $y''(x)$ is zero for constant or linear μ values, and hence $\sum_{i=1}^m w_i b_i(x) = 0$ etc., which confers this property on the matrix \mathbf{M} . Because of these two constraints, the matrix is zero for $m < 3$. The matrix is also positive semidefinite since $\hat{y}''(x)^2 > 0$ and so the quadratic form can never be negative.

The BRI fit computations are then

- Compute the matrix \mathbf{M} as described above;
- Choose a value of λ and find the values $\hat{\mu}$ that solve the linear Eq. (7);
- Use (1) to plot the predictand;
- if desired, choose the ‘best’ value of λ .

Figure 5 shows the result of applying the smoothing BRI to the carbon-12 α -emission data of DeAngelis et al. (1995). This dataset has been extensively used to

display the smoothing capabilities of MATLAB, and is available from the MATLAB file carbon12alpha.mat. It can be seen that as the value of λ decreases, the fit tends to an interpolation of the points, as one would expect.

This example shows that there is no essential problem in using BRIs in place of smoothing or penalized regression splines.

8 Conclusions

Splines and BRIs have similar errors in curve fitting, BRIs seeming from Monte Carlo results to have a slight edge. In fitting real datasets, both methodologies gave very similar fits and parameter estimates. The barycentric interpolant, unlike a spline curve, is infinitely differentiable. It is not *a priori* clear whether this is a pro or a con of BRI fits, given that statistical data may sometimes be best represented by a step or a ramp. Our Monte-Carlo fits showed that BRIs worked as well as splines for these cases, but time will show whether infinite differentiability is overall a blessing or a curse.

Practitioners doing spline fitting and fitting BRIs face similar issues, for example the optimal placement of knots or nodes respectively. Computationally, we found BRIs easier to work with than splines, but experienced users of splines may well not agree. We recommend BRIs as an equally good alternative to splines for statistical work, both as an alternative to regression splines and to smoothing splines or penalized regression splines.

Statisticians should be aware that there is an alternative methodology to splines; however, ‘the devil is in the details’, and only time and a lot of experience will reveal whether BRIs or splines are ultimately preferable.

Future work must include the provision of an R package to parallel the R *spline* package, and this is in progress; the computations in this paper were ‘prototyped’ in fortran2003 using the NAG library of numerical routines.

Acknowledgments We would like to thank two anonymous referees, whose comments have improved the clarity of this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723
- Baker R, Jackson D (2010) Inference for meta-analysis with a suspected temporal trend. *Biom J* 52(4):538–551
- Baker RD, McHale I (2012) How good is Tiger Woods? *Math Today* 48(3):124–126
- Berrut JP (1988) Rational functions for guaranteed and experimentally well-conditioned global interpolation. *Comput Math Appl* 15:1–16
- Berrut JP, Trefethen LN (2004) Barycentric Lagrange interpolation. *SIAM Rev* 46:501–517
- Berrut JP, Welscher A (2007) Fourier and barycentric formulae for equidistant Hermite trigonometric interpolation. *Appl Comput Harm Anal* 23:307–320

- DeAngelis DJ, Calarco JR, Wise JE, Emrich HJ, Neuhausen R, Weyand H (1995) Multipole strength in ^{12}C from the $(e, e'\alpha)$ reaction for momentum transfers up to 0.61 fm^{-1} . *Phys Rev C* 52(1):61–75
- de Boor C (2001) A practical guide to splines (revised edition). Springer, New York
- Bos L, De Marchi S, Hormann K, Klein G (2012) On the Lebesgue constant of barycentric rational interpolation at equidistant nodes. *Numer Math* 121:461–471
- Eilers PHC, Marx BD (1996) Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Stat Sci* 11:89–121
- Floater MS, Hormann K (2007) Barycentric rational interpolation with no poles and high rates of approximation. *Numer Math* 107(2):315331
- Güttel S, Klein G (2012) Convergence of linear barycentric rational interpolation for analytic functions. *SIAM J Numer Anal* (to appear)
- Harrell FE Jr (2001) Regression modeling strategies. Springer, Berlin
- Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558
- Jackson D, Baker RD, Bowden J (2013) A sensitivity analysis framework for the treatment effect measure used in the meta-analysis of comparative binary data from randomised controlled trials. *Stat Med* 32:931–940
- Kauermann G, Opsomer JD (2011) Data-driven selection of the spline dimension in penalized spline regression. *Biometrika* 98(1):225–230
- Klein G (2011) An extension of the Floater–Hormann family of barycentric interpolants. Report No. 2011–3. University of Fribourg Department of Mathematics
- Klein G, Berrut JP (2010) Linear barycentric rational quadrature. Report No. 2010–2. University of Fribourg Department of Mathematics
- Klein G, Berrut JP (2012) Linear rational finite differences from derivatives of barycentric rational interpolation. *Siam J Numer Anal* 50:643–656
- Möbius AF (1827) Der Barycentrische Calcul : ein neues Hilfsmittel zur analytischen Behandlung der Geometrie. http://math-doc.ujf-grenoble.fr/cgi-bin/oeitem?id=OE_MOBIUS_1_1_0
- Mullerleile P, Mullen B (2006) Sufficiency and stability of evidence for public health interventions using cumulative meta-analysis. *Am J Public Health* 96:515–522
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes, 3rd edn. Cambridge University Press, Cambridge
- Royston P, Altman DG (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 43:429–467
- Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Cambridge University Press, Cambridge
- Sapidis N, Farin G (1990) Automatic fairing algorithm for B-spline curves. *Comput Aided Des* 22:121129
- Taylor WJ (1945) Method of Lagrangian curvilinear interpolation. *J Res Nat Bur Stand* 35:151–155
- Twenge JM, Konrath S, Foster JD, Campbell WK, Bushman BJ (2008) Egos inflating over time: a cross-temporal meta-analysis of the Narcissistic Personality Inventory. *J Pers* 76:875–901
- Usmani RA (1994) Inversion of a tridiagonal Jacobi matrix. *Linear Algebra Appl* 212–213:413–414
- Wahba G (1990) Spline models for observational data. SIAM, Philadelphia
- Wegman EJ, Wright W (1983) Splines in statistics. *J Am Stat Assoc* 78:351–365