Degree Selection Methods for Curve Approximation via Bernstein Polynomials

27 January, 2021

1 Summary

In "*Degree Selection Methods for Curve Approximation via Bernstein Polynomials*" the authors propose two methods to select the degree (m) of the Bernstein polynomial (BP) basis for use in smoothing and regression problems. Choosing the appropriate degree for a BP approximation is analogous to choosing the appropriate number of knots when using splines, and/or choosing the optimal polynomial degree (assuming the knots are spaced equally, and our interest is in a compact interval). The authors propose two methods for selecting the degree of the BP basis. The first uses prior information about the probable location of inflection points, in order to derive a distribution for m. Given a some tolerance for misrepresentation, this distribution is used to compute the minimum acceptable degree. The second method uses the difference between successive BP basis coefficients, and selects m when the difference becomes small. The authors study the performance of their proposed methods in a simulation study, and by applying their methods to a well known data set of human heights.

The author's proposed methods may have some value. However, the manuscript in its current state has too many errors, and the text is at times incomprehensible. I cannot recommend the manuscript be accepted in its current state. I have included some comments, predominately on the method and simulation study, that I hope the authors find useful when revising their manuscript.

2 Major concerns

1. The authors second method is based on the degree elevation rule, discussed in Farouki (2012). However, I am not convinced that it is appropriate for this setting. Specifically, the degree elevation rule allows one to represent a *known* polynomial using a higher degree Bernstein polynomial. Note that the true basis coefficients must be known for this equality to hold. The authors are concerned with approximating an unknown curve, using noisy samples from that curve. To me, the change of setting, from curve re-expression to a finite sample of noisy realisations from said curve, fundamentally changes the problem to one where the degree elevation rule is not automatically true. Nor is it automatically sensible to inspect the difference between between $\boldsymbol{\xi}_m = (\xi_{1,m}, \dots, \xi_m)$ obtained via statistical estimation to $\tilde{\boldsymbol{\xi}}_m^r$ obtained via degree elevation from some $\boldsymbol{\xi}_r$ for $r < m$, or base degree selection methods off this difference. Say we have n data points. A a polynomial of degree n, fitted by least-squares or equivalent, is going to perfectly interpolate the data. For any $r < n$ and $r < m = n$, we can obtain $\boldsymbol{\xi}_m$ via least-squares, and $\tilde{xi}_m^r$ by elevating the degree r polynomial to one of degree m. It seems extremely unlikely that, even for very large r, we will get back the perfect interpolating polynomial from the degree elevation process. We can then extend this argument to the more relevant $m < n$ setting, where again I am not convinced that $\boldsymbol{\xi}_m$ and $\tilde{\boldsymbol{\xi}}_m^r$ are comparable. Such a claim should have some kind of associated proof, at least of asymptotic consistency $(\mathrm{E}[\tilde{xi}_m^r] = \boldsymbol{\xi}_m$ for $n \to \infty)$.

2. The authors first method requires quite precise knowledge of the location of turning points for sensible performance. I do not think I can advocate for fitting a polynomial of degree $m > 20$, so the results in Table 1 are a bit concerning. Even the lest informative, non-flat setting for U (the row obtained using Beta$(3, 3)$ random variables) implies one is quite certain that the turning point is not within the first or last 15% of the domain. More generally, if I knew the location of the turning points, I would be very close to knowing the ideal degree polynomial to use in my situation. I think my issue here is with Step 1 of the proposed process for selecting a minimum m. I cannot think what exactly would sufficiently reassure me, but some form of sensitivity analysis seems wise. Perhaps the simulation study could be extended to include some choices for U_1 and U_2 which are mis-specified in some way?

3. The simulation study is, as written, slightly nonsensical. It does not make sense to think of $i = 1, \ldots, n$ individuals with J_i data points from the same curve. This is just simulating $\sum_{i=1}^{n} J_i$ values from the true curve at points biased towards $t = 0$. Additionally, if it were actually true that one was considering n individual trajectories from slight variations of the same curve, I'm not sure that suggesting a degree 6 polynomial make sense when many of the individuals have fewer than 6 observations. There are many other inconsistencies with the simulation study as it is currently written that do not make sense, which I note below.

4. The authors focus seems to switch from choosing m to locating inflection/turning points in the fitted curve. I elaborate below, but I don't think inspecting the estimated BP coefficients is an sensible way to do this.

5. My last major concern, and perhaps it confounds the other concerns, is about the quality of the text in the manuscript. There are a number of mathematical errors and incoherent notational choices in this manuscript, and the written text is not always comprehensible. My other concerns may be alleviated if the manuscript text were significantly revised, and the authors intentions better communicated. I have made nonexhaustive notes of the errors I have spotted below.

2.1 Introduction

- This section could be much more concise, particular the 1st, 6th, and 7th paragraphs.
- It is not clear from this introduction *why* I would choose BPs over a B-spline/M-spline/I-spline. Runge's phenomenon is an okay reason to opt for something other than a polynomial, but it is not often relevant to applied problems. It would be unusual too have equally spaced interpolation points for continuous covariates in regression, or attempt density estimation of continuous distributions at equally spaced points. Additionally, polynomials are typically fit by least squares or Bayesian methods, which mitigate the impact of Runge's phenomenon.
- Page 2, line 37: Unclear who 'Guan' is.
- Page 3, line 23: "*This instruction is probabilistic method based on a previous knowledge of a possibly existing turning point on the target function.*" - I do not understand what this sentence is trying to convey.

2.2 Bernstein Polynomials to approximate curves

- The notation is inconsistent in this section. Why now talk about BPs of degree $m - 1$ (i.e. a degree 2 BP has $m = 3$) instead of m? Similarly, why is f now a function of t instead of x? Figure 1 uses x as in the introduction.
- Additionally, the notation doesn't work. Equation (2) states that the BP approximation is $\sum_{k=1}^{m} \xi_{k,m-1} b_{k,m-1} \left(\frac{t}{T_{\max}} \right)$, but in Section 1 the authors have defined $b_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$. The former permits terms such as $b_{m,m-1}$, which disagree with the notation of the latter.

- Labels of Figure 1 are all too small to read.
- Page 5, line 24: "*Therefore, the vector of Bernstein basis does not affect the form of the approximation directly.*" I do not understand what this sentence is trying to convey. The vector of basis coefficients ($\boldsymbol{\eta}$) clear affects the BP approximation, and the choice of m constrains the space of possible curves one could estimate, so it too affects the approximation (e.g.)
- Page 5, line 27: "change of behaviour" is not a precise description of what the authors are interested in, which is the location of the roots of the derivative, i.e. inflection/turning points.
  - Practically, I'm not sure why one would try to locate inflection points by inspecting the difference between successive BP coefficients. Finding the roots of the BP (And hence, by the similarity property the authors note, the roots of its derivative) is studied in Spencer (1994), who notes the advantage the BP basis has for root finding over the regular monomial basis, and the existence of efficient algorithms to identify said roots. I understand that this is not the focus of the article, and that the difference in BP coefficients is motivation to consider the distribution of M, but I wouldn't approach the problem of inflection point finding in this way.
- Page 5, line 46: The authors introduce M as a random variable, whose distribution describes the minimum m required to "*capture the change point in the interval* $(u_{(1)}, u_{(2)})$". Could the authors expand on *why* this particular random variable has the distribution of interest? It is not at all obvious to me.
- Page 7, line 13: The suggested process seems to have a circular dependency. If I suspect there is a turning point within $(a, b) \subset [0, 1]$, then isn't this process inherently more likely to produce a BP approximation that has such a turning point? For example, If I suspect there is a turning point in an interval close to either 0 or 1, then this process will select a very high degree BP, which is more flexible and thus more likely to contain a turning point in my original interval, which is likely to be spurious.
- Page 8, line 14: Is this not because the authors are trying approximate an unknown function using noisy samples, whilst 'degree elevation' is the process of re-expressing known polynomials of degree m using a BP of degree $m + r$, i.e. without noise? Why should this relationship still hold in the regression setting?
- Page 8, line 23: How large does m have to be before the estimated curve stabilises?
- Page 8, line 26: the authors state that the subsequent criteria are designed to find an *optimal* m, but the previous page states the purpose of the criteria are for finding *reasonable maximum values of* m. Perhaps clarifying the constraints will help make clear the purpose of these criteria, i.e. are the constraints computational, or issues with overfitting? Also, if the criteria are designed to find an optimal m, does that not negate role of the M-based method?
- Page 8, line 33: The notation $D^{(s)} implicitly assumes that the authors are estimating ξ via MCMC, but this is not mentioned anywhere prior.
- Page 9, line 12: If the integrate variable is $t\,/\,T_{\text{max}}$, shouldn't the integration limits be from 0 to T_{max}. Also, Should D_{m-1} use a different symbol, so as not to conflict with the previous criterion?
- Page 9, line 25: Doesn't the result in Equation (8) demonstrate that the two proposed criteria are effectively the same? One is the $\lessdot_1$ distance between the coefficients, the other, the $\lessdot_2$ distance? Both criteria are really functions of the distance between the estimated curves.
- Page 9, line 33: This does not make any sense as written, Equation (8) returns a single number and hence Median(D_{m-1}) is not a sensible quantity. I assume the authors are using the MCMC samples? This should be included as an index if so?
- In general, can the authors say anything about the performance of the Sign test / Wilcoxon Rank Test for these hypotheses? Also, doesn't using the Median in the hypothesis remove the ability to do these kinds of tests? Using the author's notation, there will be only one value of $\mathrm{Median}\,(\mathbf{D}_{m-1})$ for S MCMC samples. I don't think you can do a hypothesis test using 1 observation for each group?

2.3 Simulation Study

- Page 10, line 16: Calling n 'sample size' is unclear. There are n individual in the simulated data set, and each individual has between 3 and 10 measurements.
- Page 10, line 24: The simulation study seems fundamentally unidentifiable (as written) because noise is added twice. Which is to say, it seems that noise is added to $\boldsymbol{f}_i$ to get $\boldsymbol{W}_i$ (Page 10, line 24), and then *again* add noise in the form of $\varepsilon(t_{ij})$.
- Page 10, line 25: $\boldsymbol{f}_i$ is not defined. As far as I can tell each individual is a realisation from the *same* underlying curve f, with independent noise added. If this is the case, then it the longitudinal framework does not make sense, as all data are realisations from the same underlying functions. I suggest redesigning the simulation study to match the author's written intentions.
- Page 10, line 28: It seems unreasonable to think that it would be possible to know *"the function has a turning point at* $\frac{1}{4}$ *and* $\frac{3}{4}$*"* given some curves have only 3 observations, one of which is, by design, at $t = 0$.
- I apologise for being repetitive, but if one aim of the research is to identify inflection points, why not just look at roots of the derivative of the fitted BP? The analytic nature of the derivative of the BP is, after all, one of its best features.
  - This method would also be much more precise than trying to infer the location of inflection points by looking at the coefficients of the fitted curve. Given the author's Bayesian estimation procedure, one could also get the distribution of inflection points, and incidentally quantify the uncertainty of said inflection points. At the moment, there is no uncertainty in the estimates of the location of inflection points.
- Page 11, line 1: I find it hard to believe that a polynomial of degree 6 is appropriate for all individuals in the data when some have only 3 data points.
- Page 11, line 39: The author's previous method has produced an estimate of a minimum $m = 6$, is it necessary to consider $m \leq 6$ for the second and third methods?
- Page 11, line 40: 'lag' should be 'thinning period' (I think, not clear)
- Page 13, line 40: 'straight' should be 'solid'.
- Page 15, line 16: One could acquire a very accurate estimate of inflection point, using a lower degree polynomial, by finding the roots of the polynomial.

2.4 Berkeley Growth Study

- Page 17, line 14: Figure 4 is a plot of the *rate* of growth, i.e. the derivative of the fitted curve. It is not, as the text states, the estimated mean curve.

Misc

- The method would be accessible if code were supplied that implements the author's method, and the code for the examples in the paper was available for practitioners to experiment with.

Bibliography

Farouki, R. T. (2012), "The Bernstein polynomial basis: A centennial retrospective," *Computer Aided Geometric Design*, 29, 379–419. https://doi.org/10.1016/j.cagd.2012.03.001.

Spencer, M. R. (1994), "Polynomial Real Root Finding in Bernstein Form," PhD thesis, Brigham Young University.