Why Bernstein Polynomials: Yet Another Explanation

Olga Kosheleva and Vladik Kreinovich

Abstract In many computational situations – in particular, in computations under interval or fuzzy uncertainty – it is convenient to approximate a function by a polynomial. Usually, a polynomial is represented by coefficients at its monomials. However, in many cases, it turns out more efficient to represent a general polynomial by using a different basis – of so-called Bernstein polynomials. In this paper, we provide a new explanation for the computational efficiency of this basis.

1 Formulation of the Problem

What is a general problem. In many computational situations, it is convenient to approximate a function by a polynomial. This is, for example, how most special functions like $\sin(x)$, $\cos(x)$, and $\exp(x)$ are computed in a computer: what the computer actually computes is the sum of the first several terms in their Taylor expansion.

From the computational viewpoint, a natural question is: how can we represent a general polynomial of a given degree?

How this problem is solved in most cases. The usual way to represent a polynomial $f(x)$ of degree $\le n$ is to represent it as linear combination of corresponding monomials, i.e., as

$$f(x) = c_0 \cdot e_0(x) + c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x), \tag{1}$$

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

where c_i are arbitrary coefficients, and $e_i(x) = x^i$ for all i.

What if we process probabilities: enter Bernstein polynomials. In principle, in the linear space formed by all such polynomials, we can select a different basis $e_i(x)$. For example, in situations when we know that x can only take values from the interval $[0,1]$ – e.g., if x is a probability – it is often convenient to select a different basis $e_i(x) = x^i \cdot (1-x)^{n-i}$. Elements of this basis are known as *Bernstein polynomials*.

Empirical fact. For processing probabilities – and other values limited to the interval $[0,1]$ – many other bases were tried, but Bernstein polynomials seem to be the most computationally efficient; see, e.g., [2, 3, 4, 5, 13, 17]. In particular, in many practical problems, they are efficient in interval computations [6, 9, 10, 12] and in fuzzy computations [1, 7, 11, 15, 16, 18], when:

- *we are given* an algorithm $= f(x_1,\ldots,x_n)$ and some information about uncertainty of x_i – i.e., an interval $[\underline{x}_i, \overline{x}_i]$ or a fuzzy membership function $\mu_i(x_i)$, and
- *we want to find* the resulting uncertainty in y – i.e., an interval of possible values $y = f(x_1,\ldots,x_n)$ when each x_i is in the corresponding interval, or, correspondingly, the membership function $\mu(y)$ corresponding to y.

A natural question: what is it, what is known, and what we do in this paper. A natural question is: Why is this particular basis more efficient?

A partial answer to this question was provided in [8, 14]. In this paper, we provide another explanation for this empirical fact.

2 Analysis of the Problem

Why go beyond Taylor polynomials? In many practical situations, the usual Taylor-type polynomials, i.e.., polynomials represented by the basis $e_i(x) = x^i$, work well. However, when the input x is a probability of some event, these polynomials have a problem.

Indeed, if the event has probability x, then the opposite effect has probability $1-x$. For example, if x is the probability that team A wins a match, then (in the absence of ties) $1-x$ is the probability that the opposite team B wins this match. So, if we present the situation from the viewpoint of team B, it is reasonable to consider the new input $y = 1-x$, and – if we select Taylor-type polynomials – basis consisting of functions $e_i(x) = (1-x)^i$. However, this is a completely different basis. But whether we consider it from the viewpoint of Team A or Team B, this is the same computational problem, and it does not make sense to assume that somehow the selection of the optimal basis for this computational problem depends on which team we are more interested in.

From this viewpoint, we should select the basis which should not change if we replace x with $1-x$.

Which of such bases should we select? Bernstein polynomials have the above invariant-relative-to-replacing-x-with-$(1-x)$ property, but we can also have many different bases with this property. For example, for quadratic polynomials, we can have a basis consisting of the functions x, $1-x$, and $x \cdot (1-x)$. Which basis should we select?

Since in general, Taylor-like polynomials work well, it make sense to require that when x is small (i.e., when $1-x$ is practically equal to 1) the selected functions should be asymptotically equivalent to the usual basis.

Now, we are ready to formulate our main result.

3 Main Result

Definition 1. *By a* basis*, we mean a basis $e_0(x)$, $e_1(x)$, ..., $e_n(x)$ in the linear space of all polynomials of degree $\leq n$.*

Definition 2. *We say that a basis is* view-invariant *if it is invariant with respect to changing x to $1-x$, i.e., when the set of functions forming the basis does not change if we replace x with $1-x$:*

$$\{e_0(x), e_1(x), \ldots, e_n(x)\} = \{e_0(1-x), e_1(1-x), \ldots, e_n(1-x)\}.$$

Definition 3. *We say that the basis is* asymptotically Taylor *if for small x, each function $e_i(x)$ is asymptotically equal to x_i, i.e., that*

$$\lim_{x \to 0} \frac{e_i(x)}{x^i} = 1.$$

Proposition. *The only view-invariant asymptotically Taylor basis is the basis consisting of Bernstein polynomials.*

Proof.

1°. When $x \to 0$, each polynomial $a_0 + a_1 \cdot x + \ldots$ is asymptotically equivalent to its first non-zero term. Thus, the fact that $e_i(x)$ is asymptotically equivalent to x^i means that its first non-zero term is x^i. So, the function $e_i(x)$ has the form

$$e_i(x) = x^i + a_{i+1} \cdot x^{i+1} + \ldots$$

All these terms have x^i as one of the factors, so we can conclude that $e_i(x) = x^i \cdot P_i(x)$ for some polynomial $P_i(x) = 1 + a_{i+1} \cdot x + \ldots$ for which $P_i(0) = 1$.

2°. Due to view-invariance, a similar argument applies when we consider dependence on $1-x$. So, we can conclude that each element of the basis has the form

$(1-x)^j \cdot Q_j(x)$ for some x for which $Q_j(1) = 1$, and that among $n+1$ elements of the basis, we should have elements corresponding to all $n+1$ values $j = 0, 1, \ldots, n$.

3°. Let us see what the above two properties imply about the functions $e_i(x)$. For this purpose, let us consider these functions one by one, starting with the last one $e_n(x)$.

3.1°. Let is first consider the function $e_n(x)$. According to Part 1 of this proof, this function has the form $e_n(x) = x^n \cdot P_n(x)$. The polynomial $P_n(x)$ cannot have any non-constant terms: otherwise its product with x^n would have degree higher than n – and we consider bases in the space of all polynomials of degree $\le n$. Thus, $P_n(x) = 1$ and $e_n(x) = x^n$. From the viewpoint of dependence on $1-x$, this function tends to 1 as $1-x \to 0$ – i.e., as $x \to 1$. Thus, it corresponds to the case when $e_n(x) = (1-x)^j \cdot Q_j(x)$ with $j = 0$.

3.2°. Let us now prove, by induction over k, that all the functions $e_n(x)$, $\ldots$, $e_{n-(k-1)}(x)$ have the form $e_{n-j}(x) = x^{n-j} \cdot (1-x)^j$.

We have the base case: in Part 3.1 of this proof, we proved this statement for $k = 1$. To complete the proof by induction, we need to prove the induction step. Let us assume that the above statement holds for some value k, and let us prove that it is also true for k, i.e., let us prove that $e_{n-k}(x) = x^{n-k} \cdot (1-x)^k$.

Indeed, according to the same Part 1 of the proof, this function has the form $e_{n-k}(x) = x^{n-k} \cdot P_{n-k}(x)$, i.e., in its representation as product of irreducible polynomials, it has $n-k$ factors equal to x. On the other hand, due to Part 2 of this proof, it should also have $(1-x)^j$ as a factor.

Here, we cannot have $j < k$, since functions $e_i(x)$ with such factors $(1-x)^j$ already exist – they are $e_{n-j}(x)$, and since $n+1$ basic functions correspond to $n+1$ different value j, we cannot have two functions $e_i(x)$ corresponding to the same value j.

We also cannot have $j \ge k+1$, since then the function $e_{n-k}(x)$ should have, as factors, x^{n-k} and $(1-x)^{k+1}$, i.e., have the form $x^{n-k} \cdot (1-x)^{k+1} \cdot P(x)$ and have, thus, degree at least $n+1$ – while we only consider polynomials of degree $\le n$. Thus, the only remaining choice is $j = k$, in which case $e_{n-k}(x) = x^{n-k} \cdot (1-x)^k \cdot P(x)$ for some polynomial $P(x)$.

The polynomial $P(x)$ cannot have any non-constant terms: otherwise its product with $x^{n-k} \cdot (1-x)^k$ would have degree higher than n – and we consider bases in the space of all polynomials of degree $\le n$. Thus, $P(x)$ is a constant: $P(x) = c$ for some number c. According to Part 1 of the proof, the polynomial $(1-x)^k \cdot c$ must be equal to 1 when $x = 0$. Thus, $c = 1$ and $e_{n-k}(x) = x^{n-k} \cdot (1-x)^k$.

The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in

Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. J. Garloff, "The Bernstein algorithm", *Interval Computation*, 1993, Vol. 2, pp. 154–168.
3. J. Garloff, "The Bernstein expansion and its applications", *Journal of the American Romanian Academy*, 2003, Vol. 25–27, pp. 80–85.
4. J. Garloff and B. Graf, "Solving strict polynomial inequalities by Bernstein expansion", In: N. Munro, editor, *The Use of Symbolic Methods in Control System Analysis and Design*, volume 56 of IEE Contr. Eng., London, 1999, pp. 339–352.
5. J. Garloff and A. P. Smith, "Solution of systems of polynomial equations by using Bernstein polynomials", In: G. Alefeld, J. Rohn, S. Rump, and T. Yamamoto (eds.), *Symbolic Algebraic Methods and Verification Methods – Theory and Application*, Springer-Verlag, Wien, 2001, pp. 87–97.
6. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2012.
7. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
8. V. Kreinovich, "Theoretical explanation of Bernstein polynomials' efficiency", In: K. Cohen, N. Ernest, B. Bede, and V. Kreinovich (Eds.), *Fuzzy Information Processing 2023, Proceedings of the 2023 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2023*, Cincinnati, Ohio, May 31 – June 2, 2023, Springer Lecture Notes in Networks and Systems, 2023, Vol. 751, pp. 115–126.
9. B. J. Kubica, *Interval Methods for Solving Nonlinear Constraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.
10. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
11. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
12. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
13. P. S. V. Nataraj and M. Arounassalame, "A new subdivision algorithm for the Bernstein polynomial approach to global optimization", *International Journal of Automation and Computing*, 2007, Vol. 4, pp. 342–352.
14. N. S. Nedialkov, V. Kreinovich, and S. A. Starks, "Interval Arithmetic, Affine Arithmetic, Taylor Series Methods: Why, What Next?", *Numerical Algorithms*, 2004, Vol. 37, pp. 325–336.
15. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
16. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
17. S. Ray and P. S. V. Nataraj, "A new strategy for selecting subdivision point in the Bernstein approach to polynomial optimization", *Reliable Computing*, 2010, Vol. 14, pp. 117–137.
18. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.