# Why Bernstein Polynomials: Yet Another Explanation

Olga Kosheleva and Vladik Kreinovich

[2]Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

## 1. 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?
- The usual way to represent a polynomial $f(x)$ of degree $\le n$ is to represent it as linear combination of corresponding monomials:

$$f(x) = c_0 \cdot e_0(x) + c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x).$$

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

## 2. 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*.
- For processing probabilities – and other values limited to the interval $[0, 1]$ – many other bases have been tried.
- However, Bernstein polynomials seem to be the most computationally efficient.

# 3. Bernstein polynomials (cont-d)

- In particular, in many practical problems, they are efficient in interval computations and in fuzzy computations.
- Then, we are given an algorithm $= f(x_1, \ldots, x_n)$ and some information about uncertainty of $x_i$:
  - an interval $[\underline{x}_i, \overline{x}_i]$ or
  - a fuzzy membership function $\mu_i(x_i)$.
- We want to find the resulting uncertainty in $y$:
  - 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 $y$.

## 4. A natural question

- A natural question is: Why is this particular basis more efficient?
- A partial answer to this question was provided in our previous papers.
- In this talk, we provide another explanation for this empirical fact.

## 5. 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.

## 6. Why go beyond Taylor polynomials (cont-d)

- But whether we consider it from the viewpoint of Team A or Team B, this is the same computational problem.
- 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 $1 - x$.

## 7. Which of such bases should we select?

- Bernstein polynomials have the above invariant-relative-to-replacing-$x$-with-$(1-x)$ property.
- However, 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?
- In general, Taylor-like polynomials work well.
- So, 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.

## 8. Main Result

- By a *basis*, we mean a basis $e_0(x)$, $e_1(x)$, $\ldots$, $e_n(x)$ in the linear space of all polynomials of degree $\le n$.
- We say that a basis is *view-invariant* if the set of functions forming the basis does not change if we replace $x$ with $1-x$:

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

- 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.*

## 9. Proof

- 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 $P_i(0) = 1$.
- Due to view-invariance, a similar argument applies when we consider dependence on $1 - x$.

## 10. Proof (cont-d)

- So, we can conclude:
  - that each element of the basis has the form $(1-x)^j \cdot Q_j(x)$ for some polynomial $Q_j(x)$ 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 $j = 0, 1, \ldots, n$.
- 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 $e_n(x)$.
- Let us first consider the function $e_n(x)$.
- According to what we proved earlier, this function has the form

$$e_n(x) = x^n \cdot P_n(x).$$

# 11. Proof (cont-d)

- 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 $j = 0$.
- 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: we proved this statement for $k = 1$.
- To complete the proof by induction, we need to prove the induction step.

## 12. Proof (cont-d)

- Let us assume that the above statement holds for some value $k$.
- Let us prove that it is also true for $k+1$, i.e., let us prove that

$$e_{n-k}(x) = x^{n-k} \cdot (1-x)^k.$$

- Indeed, according to what we proved, 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 what we also proved, 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 $e_{n-j}(x)$.
- 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 $j$.

## 13. Proof (cont-d)

- We also cannot have $j \geq k+1$, 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 $\leq n$.
- 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 $\leq n$.
- Thus, $P(x)$ is a constant: $P(x) = c$ for some number $c$.

## 14. Proof (cont-d)

- According to what we proved earlier, 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.

## 15. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).