FEATURE COLUMN Monthly essays on mathematical topics

From Bézier to Bernstein

Bézier curves are ubiquitous in computer graphics. ...

Bill Casselman
University of British Columbia, Vancouver, Canada
cass at math.ubc.ca

Mail to a friend   Print this article

Introduction

Bézier curves are ubiquitous in computer graphics. They were introduced implicitly into theoretical mathematics long before computers, primarily by the French mathematician Charles Hermite and the Russian mathematician Sergei Bernstein. But it was only the work of Pierre Bézier, an employee of the automobile maker Renault, and of Paul de Casteljau, of Citroen, that made these curves familiar to graphics specialists. Recently, the polynomials defined by Bernstein have become again of interest to mathematicians.
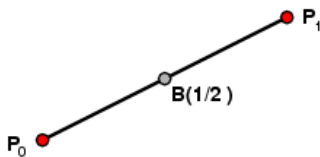
Bézier curves

The simplest Bézier curve is the straight line from one point P_0 to another P_1, with the parametric equation

$$B(t) = P_0 + t(P_1 - P_0) = (1-t) P_0 + t P_1$$

from which it follows immediately that

$$B(0) = P_0$$
$$B(1) = P_1.$$

For *t* in between *0* and *1* the point *B(t)* is *t* of the way from one to the other. This is the same as the weighted average of the two points, with P_0 given weight *1-t* and P_1 given weight *t*. When *t=1/2*, for example, *B(t)* is the point *(1/2)(P_0 + P_1)* halfway between P_0 and P_1.

A quadratic Bézier curve is determined by three control points P_0, P_1, and P_2. It has the parametric form

$$B(t) = (1-t)^2 P_0 + 2t(1-t) P_1 + t^2 P_2$$

When $t=0$ all but terms but the first vanish, and when $t=1$ all but the last vanish. Therefore
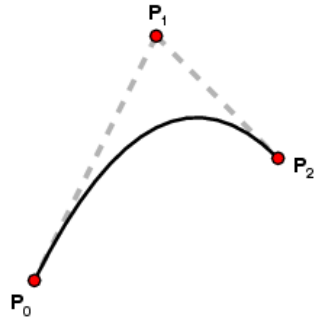
$$B(0) = P_0$$
$$B(1) = P_2.$$
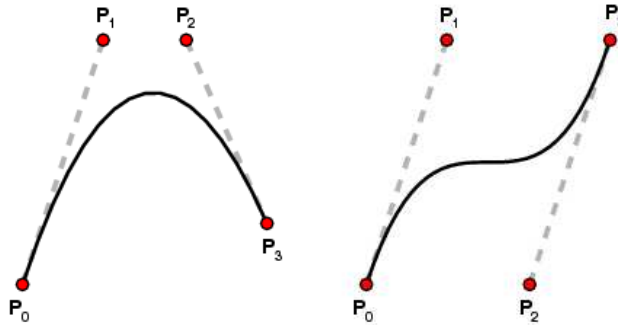
The velocity vector at t is

$$B'(t) = -2(1-t)P_0 + 2 (1-t)P_1 - 2 t P_1 + 2t P_2 = 2 [(1-t) (P_1-P_0) + t (P_2-P_1)] .$$

When $t=0$ the velocity vector is $2(P_1-P_0)$ and when $t=1$ it is $2(P_2-P_1)$. Thus, the path starts at P_0, ends at P_2, and its tangents at P_0 and P_2 intersect at P_1.

A cubic Bézier curve is determined by four control points P_0, P_1, P_2, and P_3. It has the parametric form

$$B(t) = (1-t)^3P_0 + 3 t(1-t)^2P_1 + 3t^2(1-t) P_2 + t^3P_3$$

Again, $B(0) = P_0$ and $B(1) = P_3$.

The velocity vector at t is

$$B'(t) = 3 [(1-t)^2(P_1-P_0) + 2t(1-t)(P_2-P_1) + t^2(P_3-P_2)].$$

When $t=0$ the velocity vector is $3(P_1-P_0)$ and when $t=1$ it is $3(P_3-P_2)$. Thus, the path starts at P_0, ends at P_3; when it leaves P_0 it is heading towards P_1, and when it arrives at P_3 it is coming from the direction of P_2. Otherwise, the relationship between the path and the control points is intuitively weak.

In all these cases, the coefficients of the points P_i in the parametric equation are terms which appear in the binomial expansion, respectively

$(1-t)$, t
$(1-t)^2$, $2t(1-t)$, t^2
$(1-t)^3$, $3t(1-t)^2$, $3t^2(1-t)$, t^3 .

For t between 0 and 1 these are non-negative, and by the binomial theorem the sum is $[\ t + (1\text{-}t)\ ]^n = 1^n = 1$ for $n=1$, 2, 3. What this means is that each point $B(t)$ is a weighted average of the control points, hence lies inside the convex hull of those points.
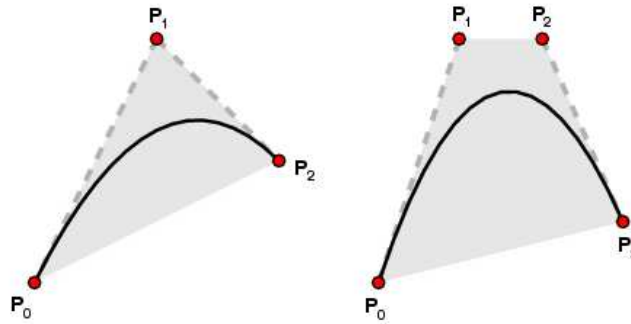
A Bézier curve of one degree can be reproduced by one of higher degree. For example, the quadratic Bézier curve with control points P_0, P_1, P_2 is the same as the cubic Bézier curve with control points

$$Q_0 = P_0$$
$$Q_1 = (1/3)P_0 + (2/3)P_1$$
$$Q_2 = (2/3)P_1 + (1/3)P_2$$
$$Q_3 = P_2$$
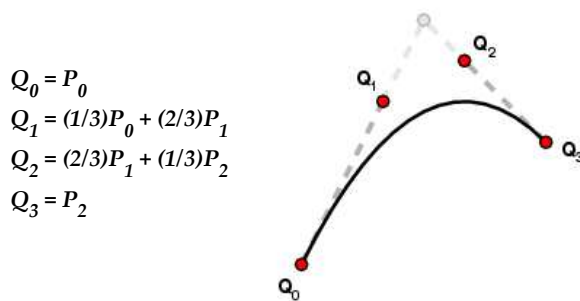
This can be checked by a simple algebraic calculation.

Calculus and Bézier curves

Suppose $B(t)$ to be a cubic Bézier curve. We know that its coordinates are functions of t, and that position and velocity at endpoints are determined by the equations:

$$B(0) = P_0$$
$$B'(0) = 3(P_1\text{-}P_0)$$
$$B'(1) = 3(P_3\text{-}P_2)$$
$$B(1) = P_3\ .$$

These conditions determine $B(t)$ uniquely, because of a simple result first observed by the French mathematician Charles Hermite. The conditions above are really conditions on the coordinates of $B(t)$, so the claim follows from this:

THEOREM.

Given numbers y_0, v_0, y_1, v_1, there exists a unique polynomial $P(x)$ of degree 3 such that

$$P(0) = y_0$$
$$P'(0) = v_0$$
$$P(1) = y_1$$
$$P'(1) = v_1\ .$$

We can guess a formula for $P(t)$, from that of Bézier cubic curves:

$P(t) = (1-t)^3 y_0 + 3(1-t)^2 t\ [y_0 + (1/3)v_0] + 3(1-t)t^2\ [\ y_1 - (1/3)v_1] + t^3\ y_1$.

As for uniqueness, if we have two such polynomials, let $D(x) = a + bx + cx^2 + dx^3$ be their difference. Then

$P(0) = a = 0$
$P'(0) = b = 0$
$P'(1) = b + 2c + 3d = 0$
$P(1) = a + b + c + d = 0$.

These tell us immediately that $a=b=0$, and these four equations immediately reduce to two:

$2c + 3d = 0$
$c + d = 0$,

which have only the solutions $c=d=0$. Q.E.D.

This is a special case of the main result of Hermite, according to which we may find a unique polynomial of degree $n-1$ satisfying a total of n conditions on values of the low order derivatives at possibly different points. This is called Hermite interpolation. This Theorem suggests a converse process. Suppose we are given a formula for a function $f(x)$, as well as a formula for its derivative $f'(x)$, and suppose we want to graph the function in a range $a \le x \le b$. We could just draw a bunch of very small line segments, but we can get a smoother curve by using Bézier cubic curves. Because if we are given a mediumly small range $[c, d]$ it seems like a reasonable idea to approximate the function over that range by a cubic polynomial $P(x)$ that satisfies these conditions:

$P(c) = f(c)$
$P'(c) = f'(c)$
$P'(d) = f'(d)$
$P(d) = f(d)$.

To see how Hermite's result applies, replace P varying over the range $[c,d]$ by $Q(t)$ where t varies over $[0,1]$, and with

$Q(t) = P((1-t)c + td),\ P(x) = Q((x-c)/(d-c))$.

To this Q we can apply Hermite's formula. A little algebra will then show that the graph between c and d will be approximated by the Bézier cubic with control points

$(c, f(c))$
$(c + h/3, f(c) + (h/3)f'(c))$
$(d - h/3, f(d) - (h/3)f'(d))$
$(d, f(d))$

where $h=d-c$.

There is one situation where this technique is definitely a good idea. Suppose that

$$f(x) = \int_0^x g(t)\,dt$$

where we do not have a formula for the indefinite integral of $g(x)$. For example, we might have

$$f(x) = \int_0^x e^{-\pi t^2}\,dt$$

Then we must approximate the integral numerically, say by Simpson's rule, so we have a recursive estimate

$$f(x+h) = f(x) + (h/6) \ (\ g(x) + 4 \ g(x + h/2) + g(x+h) \ )$$

and, by the Fundamental Theorem of Calculus, the graph between x and $x+h$ for small h is approximated by the Bézier curve with control points

$(x, f(x))$
$(x + h/3, f(x) + (h/3)g(x))$
$(x + 2h/3, f(x+h) - (h/3)g(x+h))$
$(x + h, f(x+h))$

where the last is computed according to Simpson's rule.

The same idea can be nicely used to plot the trajectories of systems of differential equations in the plane, using a numerical approximation to go from $f(x)$ to $f(x+h)$, such as one of the Runge-Kutta formulas.
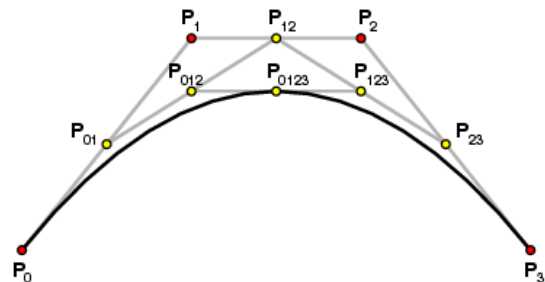
Why Bézier curves?

There is something at first a bit odd about cubic Bézier curves, the fact that the curves don't actually go through the control points. So let's imagine an alternate way to draw curves that go nicely from one point P_0 to another point P_3, with a cubic parametrization $P(t)$. The only technique that suggests itself naturally is choosing two points P_1 and P_2, and having the curve pass through them, not just near them as Bézier curves do. This means also choosing parameter values t_1 and t_2 with $P(t_1) = P_1$ and $P(t_2) = P_2$. The first choice that comes to mind is $t_1 = 1/3$, $t_2 = 2/3$. But it is well known (and explained clearly in the book by Henrici referred to later) that the curves chosen in this way look very odd. What turns out to be more reasonable is to choose t_1 closer to 0, t_2 closer to 1. But this means that we are very close to fixing values of the velocity at 0 and 1, which is essentially what Bézier curves do.
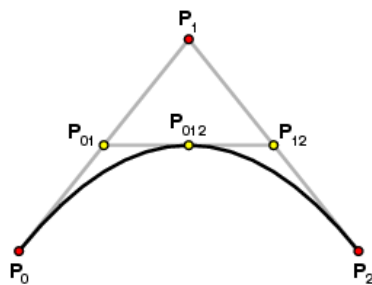
The deciding factor in choosing Bézier curves is a recursive property that makes them extremely efficient to compute. There is an algorithm attributed to de Casteljau that draws them very quickly. This depends on a certain recursive property. Suppose we are given a quadratic Bézier curve with control points P_i. Let's now divide it up into two halves, with $P_{012} = B(1/2)$. Let P_{01} be the point midway between P_0 and P_1, P_{12} be that midway between P_1 and P_2. In these circumstances: (1) *the point P_{012} is the point midway between P_{01} and P_{12}*; (2) *the half curve between P_0 and the midpoint P_{012} is again a quadratic Bézier curve with control points P_0, P_{01}, and P_{012}.* Similarly for the second half.

Therefore, in order to draw the curve the computer can keep calculating midpoints - very easy to do on modern binary machines since it involves division by 2 - until it has broken the curve into a number of small segments that are essentially straight, and then it can draw all those as straight line segments. Something very similar happens for cubic Bézier curves, too - each half of a cubic Bézier curve is a Bézier curve with easily calculated control points, as the neighboring figure illustrates.

Bézier curves and fonts

One of the most common uses of Bézier curves is in the design of fonts. Cubic Bézier curves are used in Type 1 fonts, and quadratic Bézier curves are used in True Type fonts. Cubic Bézier curves are also used in the TEX fonts designed by Donald Knuth, and one of the clearest explanations is in his book **MetaFont: the Program**.

Bernstein polynomials

The formulas for the coordinates of Bézier curves

$$B_1(t) = (1-t)\, x_0 + t\, x_1$$

$$B_2(t) = (1-t)^2\, x_0 + 2\,(1-t)t\, x_1 + t^2\, x_2$$

$$B_3(t) = (1-t)^3\, x_0 + 3\,(1-t)^2 t\, x_1 + 3(1-t)t^2\, x_2 + t^3$$

$$x_3$$

are special cases of a more general formula for polynomials of all degrees

$$B_{n,x}(t) = (1-t)^n\, x_0 + n\,(1-t)^{n-1}t\, x_1 + n(n-1)/2\,(1-t)^{n-2}t^2\, x_2 + \ldots + C_{n,k}\,(1-t)^{n-k}t^k\, x_k + \ldots + t^n\, x_n$$

where $C_{n,k}$ is the binomial coefficient $n!/k!(n-k)!$ Thus $B_{n,x}(0) = x_0$, $B_{n,x}(1) = x_n$. Here x is the array (x_i).

These polynomials were first defined by the Russian mathematician Sergei Bernstein around 1910. The coefficients of a Bernstein polynomial of degree n are the coefficients in the binomial expansion of $[(1-t) + t]^n$. They are all non-negative, and their sum is 1. The value of $B_n(t)$ is therefore a weighted average of the numbers x_i in the array x. The term $C_{n,k}\,(1-t)^{n-k}t^k$ is also the probability of k successes in n trials in which the probability of a success is t. This is not an unimportant fact, as we shall see. They have a number of fundamental properties that we have seen already for Bézier curves:

The derivative of $B_{n,x}(t)$ is $B'_{n,x}(t) = nB_{n-1,dx}(t)$ where dx is the array of differences $(x_{i+1} - x_i)$.

Thus both $B'_{n,x}(0)$ and $B'_{n,x}(1)$ have simple expressions. Otherwise, the exact relationship between the polynomial and its control values is rather obscure. One other useful feature is that

The graph of $y = B_{n,x}(t)$ for t between 0 and 1 lies inside the convex hull of the points $(k/n, x_k)$.

A related fact, which is crucial in the application of Bernstein polynomials to optimization problems:

The maximum and minimum values of $B_{n,x}(t)$ in [0,1] are bounded by the maximum and minimum values of its coefficients.

And here is the proper generalization of de Casteljau's algorithm:

The restrictions of $B_{n,x}(t)$ to each half-range [0,1/2] and [1/2,1] are also Bernstein polynomials, in the sense that each of $B_{n,x}$ (2t) and $B_{n,x}$ (1-2t) can be expressed as $B_{n,y}$ (t) for some array y that is simple to calculate.

These polynomials were invented by Sergei Bernstein in order to prove a fundamental result in approximation theory, as we shall see in the next section. They are impractical in this role, but in recent years they have proven important in optimization theory. See, for example, the thesis of Roland Zumkeller (linked to in the References). This is gratifying - almost anyone who has seen Bernstein's proof of Weierstrass' theorem must have felt that these polynomials were destined to play other roles, as well.

## Bernstein's proof of Weierstrass' approximation theorem

In about 1885 Karl Weierstrass, at the age of 70, published a proof of one of the theorems for which he is most famous: *Any continuous function on the interval [0,1] may be uniformly approximated, arbitrarily closely, by polynomials*. The definition of continuous functions is relatively simple, and this definition is in close accord with intuition, so it was apparently surprising to learn, as mathematicians of the nineteenth century acquired more and more knowledge about them, that continuous functions could exhibit rather strange behavior. Weierstrass' theorem was perhaps considered refreshing evidence that they were not so wildly behaved after all.

Many outstanding mathematicians soon came up with proofs rather different from Weierstrass' original one. One of the most satisfying was that found by Sergei Bernstein, and it was apparently for this purpose that he came up with the polynomials now named after him. Unlike other versions, his provides a very explicit converging sequence of polynomials.

Suppose f is a continuous function on *[0,1]*. For each n let f_n be the array of *n+1* values $f(k/n)$ for $k = 0, ..., n$. Explicitly, we have

$$B_{n,f}(t) = B_{n,f_n}(t) = f(0)\ (1-t)^n + f(1/n)\ n\ (1-t)^{n-1}\ t + ... + f(k/n)\ C_{n,k}\ (1-x)^{n-k}t^k + ... + f(1)\ t^n$$

THEOREM. *Suppose $f(x)$ to be a continuous function on [0,1]. The polynomials $B_{n,f}(x)$ converge uniformly on [0,1] to $f(x)$ as n goes to infinity.*
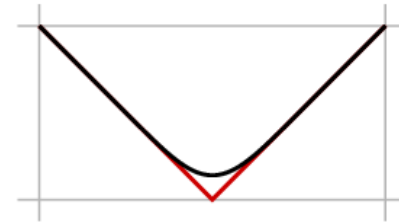
Here for example, is Bernstein's approximation to *$f(x) = |x-1/2|$* for *n=32*:

and here is a table of the Bernstein approximations of *$|t-1/2|$* at *t=1/2* for various values of *n*:

n	$B_{n,f}(1/2)$
4	*0.1367*
8	*0.0982*
16	*0.0700*
32	*0.0497*
64	*0.0352*
128	*0.0249*

As you can guess, the convergence here is of order about *$1/n^{1/2}$* around *t=1/2*, and better at other points. We can see that although one virtue of Bernstein polynomials is that they can approximate arbitrary continuous functions, on functions that one sees in practice they converge at an impractically slow rate.

Bernstein's proof of convergence is very elegant. It is a variation on a simple argument used in a slightly different context, that of **local averaging**. Suppose f to be a continuous function on *[0,1]*. Define for each n a function

$$f_n(x) = (1/2)(\ f(x - 1/2n) + f(x + 1/2n)\ )\ .$$

In other words, the value of f_n at x is the average of its values at *x-1/2n* and *x+1/2n*, an interval of width *1/n*. Or, to give slightly more complicated examples:

$$f_n(x) = (1/4)(\ f(x - 1/n) + 2\ f(x) + f(x + 1/n)\ )$$

or

$$f_n(x) = (1/8)\ (\ f(x - 3/2n) + 3\ f(x - 1/2n) + 3\ f(x + 1/2n) + f(x + 3/2n)\ )\ .$$

These are now different **weighted averages** of f near x. Bernstein's definition of $f_n(x)$ is also a weighted average of values of f around x, but *the weighting depends on x*. The value of the polynomial $B_{n,f}$ at x is also a weighted average of f around x, although that is not at all apparent at first sight. It is a sum of values of f at certain point k/n, and what happens is that *only the terms for k/n near x make a sizable contribution*. Here are some graphs of the weighting for various values of x, with $n=100$:
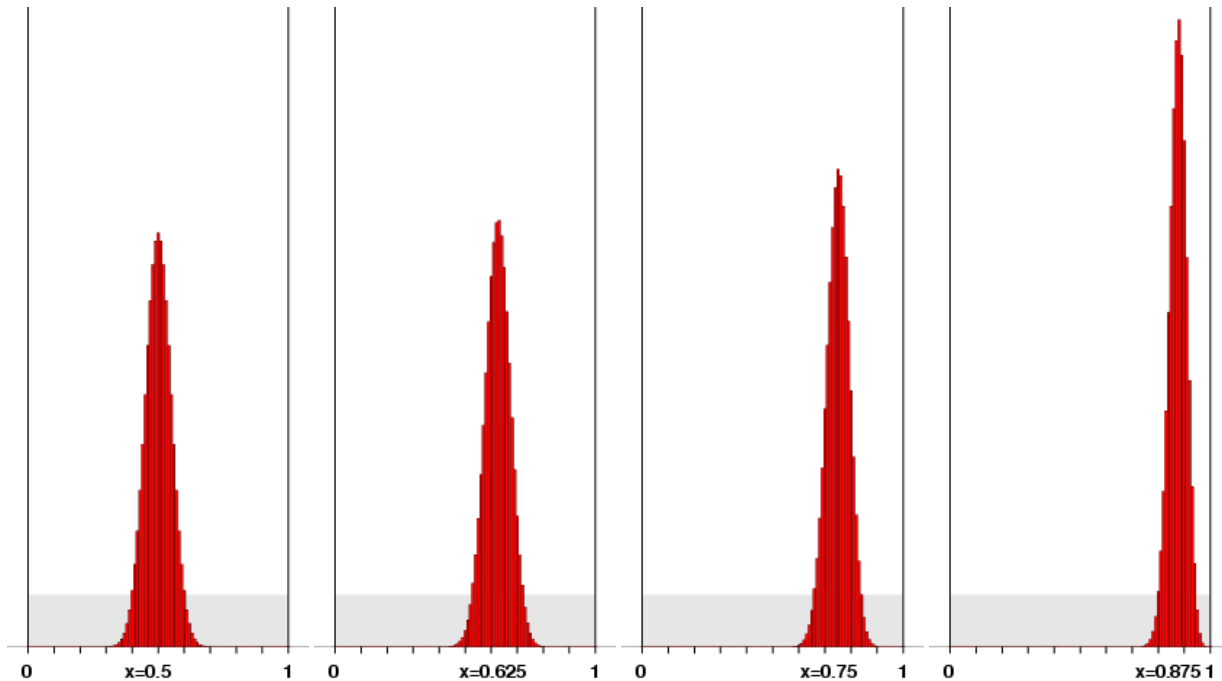
Why is this? The weighting function is the distribution of k successes in n trials with a probability of success equal to x. The average number of successes - the mean value of this probability distribution - is x, and the real basis of Bernstein's reasoning is that the standard deviation of this distribution - the quantitative measure of its spread - is

$$\sigma = \sqrt{x(1-x)/n}$$

As n grows, this shrinks. What this means is that the probability distribution clusters more narrowly around the mean value x, and more or less uniformly for all x. The proof uses this observation. Since the sum of the binomial coefficients is 1, we have

$$| f(x) - B_{n,f_n}(x) | \leq \sum_k | f(x) - f(k/n) | \, C_{n,k} \, (1-x)^{n-k} \, x^k$$

We can divide the sum over the point k/n into those which are near x and those far away, since we expect the sum over the near ones will be the major contribution. Say we choose some small number δ and sum over k/n with $|x-k/n| \leq \delta$ and those with $|x-k/n| > \delta$.

The first is

$$\sum_{|x-k/n| \leq \delta} | f(x) - f(k/n) | \, C_{n,k} \, (1-x)^{n-k} \, x^k \, .$$

If we choose δ small enough so $| f(x) - f(k/n) | < \varepsilon/2$ whenever $|x - k/n| \leq \delta$, this will be less than $\varepsilon/2$.

Now let let M be the maximum spread of f on $[0,1]$. The second sum is

$$\sum_{|x-k/n| \leq \delta} | f(x) - f(k/n) | \, C_{n,k} \, (1-x)^{n-k} \, x^k$$

which is at most M times the sum

$$\sum\nolimits_{|x-k/n| > \delta} C_{n,k} (1-x)^{n-k} x^k .$$

We must now choose n so that this sum is at most $\varepsilon/2M$. Being able to do so is plausible, since as n increases the range over which $C_{n,k}$ $(1-x)^{n-k} x^k$ is sizable becomes narrower. There are many ways to make this idea rigorous - the simplest is to use **Chebyshev's inequality**. This says that if p_i is a discrete probability distribution with mean m and standard deviation σ then the probability of $|t - m| \geq s\,\sigma$ is at most $1/s^2$. This is a very elementary fact, and we'll come back to justify it in a moment. Given it, we see that if $n > 2M\,x(1-x) / \varepsilon\,\delta^2$ then the sum above is less than $\varepsilon/2$.

Proof of Chebyshev's inequality. Since $|t - m|/s\sigma > 1$ if and only if $|t - m|^2/s^2\,\sigma^2 > 1$, the sum is that of all p_i over all t_i with $|t_i - m|^2/s^2$ $\sigma^2 \geq 1$. But this is at most the sum of $|t_i-m|^2/s^2\sigma^2$ over the same region, which is in turn is at most the sum over all i of the same terms, which is (by definition of standard deviation)

$$\sum\nolimits_i p_i\,(\,|t_i - m|^2/s^2\,\sigma^2\,) = (\,1/s^2\,\sigma^2\,)\,\sum\nolimits_i p_i\,|x_i - m|^2 = 1/s^2$$

References

To find out more ...

... about Bézier curves

The first two books contain articles by Pierre Bézier on the history of his work with the curves named after him.

- R. E. Barnhill and R. F. Riesenfeld (editors), **Computer Aided Geometric Design**, Academic Press, 1974.
  This book contains papers presented at a conference at the University of Utah that initiated much of modern computer graphics.
- G. Farin, **Curves and surfaces for computer aided design**, Academic Press, 1988.

Various works by Donald Knuth explain clearly the role of Bézier curves in font design.

- D. E. Knuth, **MetaFont: the Program**, Addison-Wesley, 1986.
  The source code for this book is in the WEB file `mf.web` available in CTAN distributions, and in particular at **http://www.ctan.org/tex-archive/systems/knuth/dist/mf/**. This can be turned to a TEX file by applying the program `weave`. Pages 123 - 131 explain extremely clearly the author's implementation of Bézier curves in his program **MetaFont**. For the admittedly rare programmer who wishes to build his own implementation (at the level of pixels), or for anyone who wants to see what attention to detail in first class work really amounts to, this is the best resource available.
- MetaFont (Wikipedia)
- MetaFont: the Book (TEX source)
  Be sure also to download the macros necessary to TEX it.

The next two items are about interpolation of various kinds.

- Charles Hermite, "Sur la formule d'interpolation de Lagrange", *Journal für die reine und angewandte mathematik*, volume 84, 1877.
  This is available from the Göttingen Digital Archive.
- Peter Henrici, **Essentials of numerical analysis**, Wiley, 1982
  Discusses interpolation and the related topic of splines very clearly.

... about Bernstein polynomials

Wikipedia has several useful entries involving Bernstein polynomials and interpolation.

- Sergei Bernstein
- Bernstein polynomials
- de Casteljau's algorithm
- Runge's phenomenon

There is a major site at Technion University concerned with approximation theory:

- The entry on Sergei Bernstein
- *Demonstration of Weierstrass' approximation theorem ...*
  This is a reproduction of Bernstein's original publication in French, which is not otherwise easy to find.
- *Weierstrass and approximation theory*
  This is an historical essay by Allan Pinkus.

Bernstein polynomials are not practical for polynomial approximation, but in recent years it has become apparent that they are practical for optimization. The next item is one of the more interesting places where this is discussed.

- Roland Zumkeller's web pages
  Zumkeller's thesis *Global optimization in type theory* is available in the publications list. It explains the role of Bernstein polynomials in a project to provide provable estimates of the minimal value of a polynomial on certain domains. This is part of a larger project to produce a formal proof of of Kepler's conjecture along the lines of Thomas Hales' proof. His exposition of the theory of Bernstein polynomials is novel.

Bill Casselman
University of British Columbia, Vancouver, Canada
cass at math.ubc.ca

Those who can access JSTOR can find some of the papers mentioned above there. For those with access, the American Mathematical Society's MathSciNet can be used to get additional bibliographic information and reviews of some these materials. Some of the items above can be accessed via the ACM Portal , which also provides bibliographic services.