Steven R. Dunbar
Department of Mathematics
203 Avery Hall
University of Nebraska-Lincoln
Lincoln, NE 68588-0130
http://www.math.unl.edu
Voice: 402-472-3731
Fax: 402-472-8466

Selected Topics in Probability and Stochastic Processes
Steve Dunbar

Question of the Day

How can we approximate a continuous function with a polynomial? How do we measure the quality of the approximation? If the function is differentiable, do the derivatives of the polynomials also approximate the derivatives of the function?

Key Concepts

1. Weierstrass approximation theorem
2. Bernstein polynomials $b_n(x) = \binom{n}{j} x^j (1-x)^{n-j}$
3. Chebyshev's inequality

Vocabulary

1. Bernstein polynomials $b_n(x) = \binom{n}{j} x^j (1-x)^{n-j}$
2. Chebyshev's inequality

Mathematical Ideas

Bernstein Polynomials

The Bernstein polynomials are defined as

$$b_{n,j} \binom{n}{j} x^j (1-x)^{n-j}$$

for $j = 0, 1, \ldots, n$.

Weierstrass Approximation Theorem

This section is adapted from: "Bernstein Polynomials and Brownian Motion", by Emmanuel Kowalski, American Mathematical Monthly, December 2006, pages 865-886.

Theorem 1 (Weierstrass Approximation Theorem) Let d be a positive integer. If $f : [0,1]^d \to \mathbb{C}$ is a continuous function and if

$$B_n[f](x) = \sum_{0 \le i_1 \ldots i_d \le n} f\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{k=1}^{n} \binom{n}{i_k} x_k^{i_k} (1-x_k)^{n-i_k}$$

for $n = 1, 2, \ldots$ and $\mathbf{x} = (x_1, \ldots, x_n)$ in $[0,1]^d$, then $B_n[f] \to f$ uniformly on $[0,1]^d$.

The Big Idea of the proof is the following: Let X_{ij}, for $1 \le \le d$ and integers $j \ge 1$ be a Bernoulli random variable so that $\mathbb{P}X_{ij} = 1 = x_i$ and $\mathbb{P}X_{ij} = 0 = 1 - x_j$. Let $S_{in} = \sum_{j=1}^{n} X_{ij}$ Let $\mathbf{S}_n = (S_1, \ldots, S_n)$. By either the Weak Law or the Strong Law of Large Numbers we can assert that $\mathbf{S_n}/n \approx \mathbf{x}$. By continuity, $f(\mathbf{S}_n/n) \approx f(\mathbf{x})$. Then all the more so $\mathbb{E}\mathbf{S}_n/n \approx f(\mathbf{x})$. But this is $B_n[f](x) \approx f(x)$, and we can see that $B_n[f](x)$ is a polynomial.

Proof: Consider a positive integer n and $\mathbf{x} = (x_1, \ldots, x_d)$ in $[0,1]^d$. Let $\mathbf{X}_i = (X_{1i}, \ldots X_{di})$ be a finite sequence of random vectors defined on some probability space such that X_{ji} are all independent and each takes values 0 and 1 according to a Bernoulli distribution for which $\mathbb{P}X_{ji} = 1 = x_j$ and $\mathbb{P}X_{ji} = 0 = 1 - x_j$. (Note that the first index j is the coordinate and the second index$i$ is the sequence index.) Write $S_{jn} = \sum_{i=1}^{n} X_{ji}$ and $\mathbf{S}_n = (S_{1n}, \ldots, S_{dn})$. Then as binomial random variable,

$$\mathbb{P}S_{jn}/n = i_j/n = \binom{n}{i_j} x_j^{i_j} (1-x_j)^{n-i_j}$$

Then

$$B_n[f](x) \sum_{0 \le i_1 \ldots i_d \le n} f\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{k=1}^{n} \binom{n}{i_k} x_k^{i_k}(1-x_k)^{n-i_k}$$

noting the use of the independence and the definition of the expectation. Note also that

$$\sigma_{j,n}^2 = \operatorname{Var} S_{jn}/n = \frac{x_j(1-x_j)}{n} \le \frac{1/4n}{.}$$

(Note that the source article has a typo here.) Note that also $\mathbb{E}\mathbf{S}_n = n\mathbf{x}$, whence

$$B_n[f](\mathbf{x}) - f(\mathbf{x}) = \mathbb{E}f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{x}) = \mathbb{E}f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbb{E}\frac{\mathbf{S}_n}{n}).$$

Set $\|\mathbf{x}\| = \max_j\{|x_j|\}$ for $\mathbf{x} \in \mathbb{R}^d$. The event that the maximum of the coordinates is greater than a value is contained in the union of events where each coordinate is greater than the value. Then by Chebyshev's inequality applied component-wise, and independence, we obtain that for any $\delta > 0$ the upper bound

$$\mathbb{P}\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| \ge \delta \le \sum_{1 \le j \le d} \mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| \ge \delta \le \frac{d\sigma_{jn}^2}{\delta^2} \le \frac{d}{4n\delta^2}.$$

Jamie Radcliffe also pointed out that

$$\begin{aligned}
\mathbb{P}\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| \ge \delta &= 1 - \mathbb{P}\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| < \delta \\
&= 1 - \left(\prod_{1 \le j \le d} \mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| < \delta\right) \\
&= 1 - \left(\prod_{1 \le j \le d} \left(1 - \mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| \ge \delta\right)\right) \\
&= \sum_{1 \le j \le d} \mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| \ge \delta - \sum_{1 \le j,k \le d} \left(\mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| \ge \delta\right)\left(\mathbb{P}\left|\frac{S_{kn}}{n} - x_k\right| \ge \delta\right) + \ldots \\
&\le \sum_{1 \le j \le d} \mathbb{P}\left|\frac{S_{jn}}{n} - x_j\right| \ge \delta
\end{aligned}$$

This gives a sharper estimate.

Now let $\epsilon > 0$ be given. By the uniform continuity of f on $[0,1]^d$, there is a positive $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| < \epsilon$, whenever $\|\mathbf{x} - \mathbf{y}\| < \delta$. Then

$$\int_{\|\mathbf{S}_n/n - \mathbf{x}\| < \delta} \left| f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{x}) \right| d\mathbb{P} \le \epsilon.$$

(This is not sophisticated, we are integrating something of magnitude at most ϵ over a measure of at most 1.) Then on the complementary event, using the triangle inequality and the previous estimate of the probability of the complementary event:

$$\begin{aligned}\int_{\|\mathbf{S}_n/n - \mathbf{x}\| \ge \delta} \left| f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{x}) \right| d\mathbb{P} &\le 2 \max_{\mathbf{x}\in[0,1]^d} |f(\mathbf{x})| \mathbb{P}\left\| \frac{\mathbf{S_n}}{n} - \mathbf{x} \right\| \ge \delta \\ &\le \frac{2d}{4n\delta^2} \max_{\mathbf{x}\in[0,1]^d} |f(\mathbf{x})|.\end{aligned}$$

Putting together the integrals over the two complementary events, for sufficiently large values of n,

$$\left| \mathbb{E} f\left(\frac{\mathbf{S}_n}{n}\right) - f\left(\mathbb{E}\frac{\mathbf{S}_n}{n}\right) \right| \le 2\epsilon$$

which holds independently of n, and the theorem is established.

Lemma 1 If $f : [0,1] \to \mathbb{C}$ is a continuous function and $B_n[f]$ is the nth Bernstein polynomial of f, then

$$B_n[f]'(x) = \sum_{j=0}^{n-1} \binom{n-1}{j} \left(f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \right) x^j (1-x)^{n-1-j}$$

Proof: The proof is purely computational:

$$\begin{aligned}B_n[f]'(x) &= \sum_{j=0}^{n} \binom{n}{j} \left(j x^{j-1}(1-x)^{n-j} - (n-j) x^j (1-x)^{n-j-1} \right) \\ &= \sum_{j=1}^{n} j \binom{n}{j} x^{j-1}(1-x)^{n-j} - \sum_{j=0}^{n-1} (n-j) \binom{n}{j} x^j (1-x)^{n-j-1}\end{aligned}$$

Note the slight shift in indices on the summand in the last line, reflecting that the end terms disappear. Now making the change of summation $j = k - 1$ in the first summand and then relabeling $k = j$, and noting that

$$j\binom{n}{j} = n\binom{n-1}{j-1}$$

and

$$(n-j)\binom{n}{j} = n\binom{n}{j-1}$$

we regroup to get

$$B_n[f]'(x) = \sum_{j=0}^{n-1} \binom{n-1}{j} \left(f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right)\right) x^j (1-x)^{n-1-j}$$

Lemma 2 If $f : [0,1]^2 \to \mathbb{C}$ is a continuous function and $B_n[f]$ is the nth Bernstein polynomial of f, then

$$\partial_x B_n[f](x,y) = \sum_{j=0}^{n-1}\sum_{k=0} n\binom{n-1}{j}\binom{n}{k} n \left(f\left(\frac{j+1}{n}, \frac{k}{n}\right) - f\left(\frac{j}{n}, \frac{k}{n}\right)\right) x^j(1-x)^{n-1-j} y^k (1-y)^{n-}$$

Proof: In the two-variable Bernstein polynomial for f, grouping all the terms with $x = j/n$ and factoring we can see that

$$B_n[f](x,y) = \sum_{j=0} n\binom{n}{j} G_n\left(\frac{j}{n}, y\right) x^j (1-x)^{n-j}$$

where $G_n(x,y) = B_n[g](y)$ the Bernstein polynomial of the function of y given by $g(y) = f(x,y)$ with x fixed. Now applying the previous lemma in the one-dimensional case,

$$\partial_x B_n[f](x,y) = \sum_{j=0}^{n-1} \binom{n-1}{j} n \left(G_n\left(\frac{j+1}{n}, y\right) - G_n\left(\frac{j}{n}, y\right)\right) x^j (1-x)^{n-j}.$$

Now expand, and unfactor, and regroup to get the result of the lemma.

Theorem 2 If $f : [0, 1] \to \mathbb{C}$ is a continuous function and x_0 is a point of $[0, 1]$ such that the derivative $f'(x_0)$ exists then

$$\lim_{n\to\infty} B_n[f](x)\ f'(x)$$

Proof: The case $x_0 = 0$ follows immediately from evaluating

$$B_n[f]'(x) = \sum_{j=0}^{n-1} \binom{n-1}{j} \left(f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right)\right) x^j (1-x)^{n-1-j}$$

at $x_0 = 0$ to obtain only the $j = 0$ summand, leading to

$$B_n[f]'(0) = \frac{f(1/n) - f(0)}{1/n}.$$

Likewise $x_0 = 1$ leads to

$$B_n[f]'(1) = \frac{f(1) - f(1 - 1/n)}{1/n}.$$

So now assume that $x_0(1 - x_0) \neq 0$. For the binomial random variables, this means that $\operatorname{Var} S_n = nx_0(1 - x_0) \neq 0$.

Writing $f(x) = f(x_0) + (x - x_0) * f'(x_0) + g(x)$ with $g(x_0)$ differentiable at $x = x_0$ and $g'(x_0) = 0$. Since the operation of finding the Bernstein polynomial is linear and the derivative is linear, it is enough to show that $B_n[g]'(x_0) \to 0$ as $n \to \infty$. We can write $g(x) = (x - x_0)\epsilon(x)$ where $\lim_{x\to x_0} \epsilon(x) = 0$.

We claim that the following probabilistic formula holds, with $S_n = \sum_{j=0}^{n} X_j$ is the sum of Bernoulli random variables as before with $\mathbb{P}X_j = 1 = x_0$ and $T_n = S_n/n$ is the sequence of empirical averages:

$$B_n[g]'(x) = \frac{\mathbb{E}(T_n - x)g(T_N)}{V(T_n)}.$$

Proof of claim: Go back to the derivative expression for the Bernstein approximation, and apply it to g:

$$\begin{aligned} B_n[g]'(x) &= \frac{1}{x(1-x)} \sum_{j=0}^{n} \binom{n}{j} g\left(\frac{j}{n}\right)(j - nx)x^j(1-x)^{n-j} \\ &= \frac{n}{x(1-x)} \sum_{j=0}^{n} \binom{n}{j} g\left(\frac{j}{n}\right)(j/n - x)x^j(1-x)^{n-j}. \end{aligned}$$

After expanding

$$g\left(\frac{j}{n}\right)\left(\frac{j}{n}-x\right)=\frac{j}{n}g\left(\frac{j}{n}\right)-xg\left(\frac{j}{n}\right)$$

we recognize the expression

$$B_n[g]'(x)=\frac{n}{x(1-x)}\left(B_n[g_1](x)-xB_n[g_2](x)\right)$$

in which the Bernstein polynomials of $g_1(y)=yg(y)$ and $g_2(y)=xg(y)$ occur. Writing both of these functions in expectation Bernstein polynomial form, $B_n[g_1](x)=\mathbb{E}g_1(S_n/n)=(S_n/n)\cdot g(S_n/n)=T_n\cdot g(T_n)$, and $B_n[g_2](x)=\mathbb{E}xg(S_n/n)=\mathbb{E}xg(T_n)$, using the linearity of the expectation, and $\operatorname{Var}T_n=x(1-x)/n$ and the claim is established.

Whence for $x=x_0$,

$$B_n[g]'(x)=\frac{\mathbb{E}(T_n-x)g(T_N)}{V(T_n)}=\frac{\mathbb{E}(T_n-x)^2\epsilon(T_N)}{V(T_n)}$$

Now we use this to prove the Theorem. Let$\eta>0$ be given. There exists a $\delta>0$ such that when $|x-x_0|<\delta$ we have $|\epsilon(x)|\eta$. As in the proof of the Weierstrass Approximation Theorem, we estimate the right hand side of

$$\frac{\mathbb{E}(T_n-x)^2\epsilon(T_N)}{V(T_n)}$$

separately on the events $\{|T_n-x_0|<\delta\}$ and $\{|T_n-x_0|\geq\delta\}$. On the first event,

$$\begin{aligned}\left|\int_{\{|T_n-x_0|<\delta\}}\epsilon(T_n)(T_n-x_0)^2d\P\right|&\leq\eta\left|\int_{\{|T_n-x_0|<\delta\}}(T_n-x_0)^2\right|d\P\\&\leq\eta\left|\int_{\Omega}(T_n-x_0)^2\right|d\P\\&=\eta\operatorname{Var}T_n.\end{aligned}$$

On the second event, write $M=\max|\epsilon(x)|$and invoke Cauchy's Inequality to conclude:

$$\left|\int_{\{|T_n-x_0|\geq\delta\}}\epsilon(T_n)(T_n-x_0)^2d\P\right|^2\leq M^2\mathbb{P}|T_n-x_0|\geq\delta\mathbb{E}(T_n-x_0)^4$$

Now we can use the so-called Cantelli estimate on the fourth moment (proof given elsewhere!) $\mathbb{E}(T_n - x_0)^4 \le 1/n^2$, valid for all $x_0 \in [0,1]$. Now applying Chebyshev's inequality to the probability expression in the middle, we get:

$$\mathbb{P}|T_n - x_0| \ge \delta \le \frac{\operatorname{Var}(T_n - x_0)}{\delta^2} = \frac{x_0(1-x_0)}{n\delta^2}.$$

Using this in the estimate, and taking the square root:

$$\left| \int_{\{|T_n - x_0| \ge \delta\}} \epsilon(T_n)(T_n - x_0)^2 d\P \right| \le M \sqrt{\frac{x_0(1-x_0)}{n\delta^2}} \frac{1}{n}.$$

Then

$$|B_n[g]'(x)| \le \eta \frac{x(1-x)}{n} + M \sqrt{\frac{x_0(1-x_0)}{n\delta^2}} \frac{1}{n}.$$

This term approaches 0 as $n \to \infty$, as required.

Problems to Work for Understanding

1. Show that $x_0 = 1$ leads to

$$B_n[f]'(1) = \frac{f(1) - f(1 - 1/n)}{1/n}.$$

Reading Suggestion:

1. "Bernstein Polynomials and Brownian Motion", by Emmanuel Kowalski, American Mathematical Monthly, December 2006, pages 865-886.

Outside Readings and Links:

1.
2.
3.
4.