Do Estimates of an Integral Really Improve as *n* Increases?

SHERMAN K. STEIN
University of California
Davis, CA 95616-8633

One way to estimate $\int_0^1 f(x)\,dx$ is to partition the interval $[0,1]$ into n sections of equal lengths and use their right-hand endpoints to form the sum $\sum_{i=1}^{n} f(i/n)(1/n)$. Call this estimate E_n. If f is continuous, then E_n approaches the integral as n approaches infinity. A beginning calculus student or the author of a calculus text might ask, "As n increases, does the estimate get steadily closer to the integral?" In other words, is the approach monotonic?

A quick sketch can produce an example where E_1 equals the integral but E_2 does not. So the answer is no. However, if the derivative f' is nonnegative throughout the interval, then $E_1 \geq E_2 \geq \int_0^1 f(x)\,dx$ as another quick sketch will show. A similar conclusion holds if f' is nonpositive, in which case we have $E_1 \leq E_2 \leq \int_0^1 f(x)\,dx$. In short, if f' does not change sign in the interval $[0,1]$, then E_2 is at least as good an estimate as E_1, and similarly for any positive integers p and q, E_{pq} is at least as good as E_p.

Keeping the assumption that $f' \geq 0$, we next ask, "Is $E_2 \geq E_3$?" Once again, a diagram will show that the answer is no. And this automatically raises the next question: "For which general f with nonnegative first derivative, do we have $E_2 \geq E_3$?" At this point I talked to numerical analysts and looked in numerical analysis texts, but found no mention of the problem. Only after I had shown that $E_n \geq E_{n+1}$ for all n did I find that several mathematicians had studied this and related problems as early as 1929. I will first describe my own experience with the problem and its generalization and then the various techniques that others had employed in their solutions. Some of the arguments are elementary and could be sketched in a few minutes in a calculus class.

1. My Bout with the Problem

I began by using algebra, looking at the inequality $E_2 \geq E_3$, which, written out in full, is

$$\tfrac{1}{2}\left(f(1) + f\left(\tfrac{1}{2}\right)\right) \geq \tfrac{1}{3}\left(f(1) + f\left(\tfrac{2}{3}\right) + f\left(\tfrac{1}{3}\right)\right)$$

or, equivalently,

$$f(1) - 2f\left(\tfrac{2}{3}\right) + 3f\left(\tfrac{1}{2}\right) - 2f\left(\tfrac{1}{3}\right) \geq 0. \tag{1}$$

To avoid fractions, I replaced f by the function g defined by $g(x) = f(x/6)$, with domain $[0,6]$. Inequality (1) then read

$$g(6) - 2g(4) + 3g(3) - 2g(2) \geq 0. \tag{2}$$

What assumption about g in addition to the fact that g' is nonnegative, will imply that (2) holds?

To answer this, I used second differences. If g is a function and a and h are numbers, the expression $g(a+h)-g(a)$ is called a "first difference." Throughout h will be positive. The first difference of the first difference is called a "second difference," namely

$$[g(a+h+h)-g(a+h)]-[g(a+h)-g(a)]$$
$$=g(a+2h)-2g(a+h)+g(a).$$

The third difference, fourth difference, and so on are defined inductively, each being the first difference of the preceding one. The coefficients in the kth difference are the coefficients of the polynomial obtained by expanding $(x-1)^k$. In particular, the coefficient of the top term, $g(a+kh)$, is 1 and the coefficient of $g(a)$ is $(-1)^k$. An inductive argument shows that if g has a kth derivative then the kth difference equals $g^{(k)}(c)h^k$ for some number c in $[a, a+kh]$.

The expression "$g(6)-2g(4)$" in (2) suggested the second difference, $g(6)-2g(4)+g(2)$. If g'' is nonnegative, so is that second difference. So rewriting (2) in the form

$$[g(6)-2g(4)+g(2)]+3[g(3)-g(2)]\geq 0$$

showed that if both g' and g'' are nonnegative, then $E_2 \geq E_3$. I recorded this as a theorem.

THEOREM 1.1. *If f' and f'' are nonnegative in* $[0,1]$, *then* $E_2 \geq E_3$.

In a similar manner, with $g(x)=f(x/12)$, I found that the inequality $E_3 \geq E_4$ reduced to

$$g(12)-3g(9)+4g(8)-3g(6)+4g(4)-3g(3)\geq 0.$$

The expression "$g(12)-3g(9)$" suggested the third difference,

$$g(12)-3g(9)+3g(6)-g(3),$$

and the inequality $E_3 \geq E_4$ became

$$[g(12)-3g(9)+3g(6)-g(3)]+4g(8)-6g(6)+4g(4)+2g(3)\geq 0,$$

and then

$$[g(12)-3g(9)+3g(6)-g(3)]+4[g(8)-2g(6)+g(4)]$$
$$+2[g(16)-g(3)]\geq 0.$$

So I had Theorem 1.2.

THEOREM 1.2. *If f', f'', and f''' are nonnegative in* $[0,1]$, *then* $E_3 \geq E_4$.

These results already suggested a conjecture and an approach to proving it. In any case, I looked at the next inequality, $E_4 \geq E_5$, which is equivalent to

$$g(20)-4g(16)+5g(15)-4g(12)+5g(10)-4g(8)+5g(5)-4g(4)\geq 0. \tag{3}$$

So I started with a fourth difference, and after some straightforward arithmetic found that (3) could be written in terms of differences:

$$[g(20)-4g(16)+6g(12)-4g(8)+g(4)]+5[g(15)-2g(12)+g(9)]$$
$$+5[g(10)-g(9)]+5[g(5)-g(4)]\geq 0.$$

So I had justified the next theorem.

THEOREM 1.3. *If* f', f'', *and* $f^{(4)}$ *are nonnegative in* $[0,1]$, *then* $E_4 \geq E_5$.

The pattern was not what I expected: $f^{(3)}$ was missing. To get more information, I went on, using the same approach, obtaining the next two theorems.

THEOREM 1.4. *If* $f', f'', f^{(3)}$, *and* $f^{(5)}$ *are nonnegative in* $[0,1]$, *then* $E_5 \geq E_6$.

THEOREM 1.5. *If* $f', f'', f^{(3)}$ *and* $f^{(6)}$ *are nonnegative in* $[0,1]$, *then* $E_6 \geq E_7$.

These cases suggested that the hypotheses were very sensitive to n, but gave no hint of the general theorem. So I attacked the case $n = 7$, that is, $E_7 \geq E_8$. However, after several attempts, starting with a seventh difference, I could not express the inequality in terms of differences that met this critical condition: The lead coefficient of each difference is positive.

What to do? I felt that information about the higher derivatives must continue to imply the general inequality $E_n \geq E_{n+1}$. Moreover, it was clear that the problem was algebraic: how to represent vectors as linear combinations of certain prescribed vectors, namely, the differences, with nonnegative coefficients.

As I looked over the settled cases, I saw that f' and f'' appeared in all of them. Could it be that $f^{(3)}$, $f^{(4)}$, $f^{(5)}$, and $f^{(6)}$ were just superfluous distractions. I had been using the greedy algorithm, where a high-order difference mimicked the first two terms of the typical inequality. Maybe I should switch to a timid approach. However, I knew that I couldn't get by with just first differences, for, if I could, then that would show that if f' were nonnegative, then $E_2 \geq E_3$. So I would have to exploit at least second differences. The first test case was the inequality $E_3 \geq E_4$, which amounted to

$$g(12) - 3g(9) + 4g(8) - 3g(6) + 4g(4) - 3g(3) \geq 0. \tag{4}$$

Mimicking only $g(12)$ and a bit of the next summand, I started by writing (4) as

$$[g(12) - 2g(9) + g(6)] - g(9) + 4g(8) - 4g(6) + 4g(4) - 3g(3) \geq 0. \tag{5}$$

But $g(9)$ has the negative coefficient -1. The next second difference to use in rewriting (5) would be $-[g(9) - 2g(8) + g(7)]$, which would be nonpositive if f'' were nonnegative. That would be useless. So I had to settle the question, "Can the left side of (4) be expressed as the sum of first and second differences with positive coefficients?" The emphasis is on the world "positive".

Maybe my timid algorithm wasn't timid enough. Perhaps I had tried to do too much by mimicking $g(12)$ and part of the next term, $-3g(9)$. I wondered whether I could avoid negative coefficients by being less demanding.

I could start instead with $g(12) - 2g(10.5) + g(9)$, which would then leave me with $2g(10.5)$ as the first remaining term to mimic. However, the noninteger argument 10.5 would open a can of worms and went against all my instincts. Surely, if (4) could be represented the way I wanted, it should not require noninteger arguments.

So, instead, I took a shot in the dark, beginning with the second difference $g(12) - 2g(11) + g(10)$. It was very timid, since I didn't try to go after $g(9)$. However, it was daring, since I was introducing $g(11)$ and $g(10)$, which weren't present in (4) and therefore made me a bit uneasy. However, it did give a positive coefficient as the next term to be mimicked since (4) had become

$$\begin{aligned}[g(12) - 2g(11) + g(10)] + 2g(11) - g(10) - 3g(9) + 4g(8)\\ - 3g(6) + 4g(4) - 3g(3) \geq 0.\end{aligned} \tag{6}$$

Mimicking $2g(11)$, I used the timid $2g(11)-4g(10)+2g(9)$. So (4) now had become

$$[g(12)-2g(11)+g(10)]+2[g(11)-2g(10)+g(9)]+3g(10)-5g(9)$$
$$+4g(8)-3g(6)+4g(4)-3g(3)\geq 0.$$

The term $3g(10)$, again happily with a positive coefficient, suggested using $3[g(10)-2g(9)+g(8)]$, and I operated on (6) to get

$$\begin{aligned}&[g(12)-2g(11)+g(10)]+2[g(11)-2g(10)+g(9)]\\&\quad+3[g(10)-2g(9)+g(8)]\\&\quad+g(9)+g(8)-3g(6)+4g(4)-3g(3)\geq 0.\end{aligned}$$

Since $g(9)$ has a positive coefficient, I could use $g(9)-2g(8)+g(7)$. Two additional steps then yielded

$$\begin{aligned}&[g(12)-2g(11)+g(10)]+2[g(11)-2g(10)+g(9)]\\&\quad+3[g(10)-2g(9)+g(8)]\\&\quad+[g(9)-2g(8)+g(7)]\\&\quad+3[g(8)-2g(7)+g(6)]+5[g(7)-2g(6)+g(5)]\\&\quad+4[g(6)-2g(5)+g(4)]+3[g(5)-g(3)]\geq 0.\end{aligned}\tag{7}$$

So I had expressed (4) in terms of first and second differences, all with positive coefficients. As a consequence, I saw that Theorem 1.2 was true without any mention of $f^{(3)}$.

This timid approach was as timid as could be. At each step the argument went down only by 1, with a second difference starting at $g(12)$, then at $g(11)$, then at $g(10)$, and so on. This approach was an automatic procedure, an algorithm. All would be settled if the algorithm gave only positive coefficients for the second differences (and for the first differences that appear at the end).

I tested the algorithm for $n=4,5,6$, and the pesky case, 7. It went through smoothly, and suggested that the theorem to prove had a fixed hypothesis: If f' and f'' are nonnegative in $[0,1]$, then $E_n\geq E_{n+1}$.

I didn't bother to check the case $n=8$. The pattern was so uniform that all that remained was to describe it. The proof would then follow easily.

As I looked back at the calculations one thing struck me. The function g played no role in the manipulation. In that case it should be possible to replace differences by the polynomials they resemble. This suggested that $g(a+2h)-2g(a+h)+g(a)$ could be represented by $x^{a+2h}-2x^{a+h}+x^a=x^a(x^h-1)^2$. In that case, $c[g(a+2h)-2g(a+h)+g(a)]$ would correspond to $cx^a(x^h-1)^2$. For instance, the identity

$$g(6)-2g(4)+3g(3)-2g(2)=[g(6)-2g(4)+g(2)]+3[g(3)-g(2)]$$

reads, in the language of polynomials,

$$x^6-2x^4+3x^3-2x^2=[x^6-2x^4+x^2]+3[x^3-x^2]=x^2(x^2-1)^2+3x^2(x-1).$$

Since polynomials form a structure with addition and multiplication, they might provide a convenient arena in which to carry out the bookkeeping. This approach did work out and gave a much more direct proof of the following theorem.

THEOREM 1.6. *Let the function f be defined on* $[0,1]$ *and be twice differentiable. If* f' *and* f'' *are nonnegative, then* $E_n \geq E_{n+1}$.

In outline, the proof goes like this. First note that a second difference with $h > 1$ can be expressed as the sum of second differences with $h = 1$, all with the same sign. For instance,

$$g(6) - 2g(4) + g(2) = [g(6) - 2g(5) + g(4)] + 2[g(5) - 2g(4) + g(3)] + [g(4) - 2g(3) + g(2)].$$

To see that this holds in general, note that

$$cx^2(x^h - 1)^2 = cx^a(x^{h-1} + \cdots + 1)^2(x-1)^2.$$

Similarly, we may restrict h to 1 in our first differences.

The expression $E_n - E_{n+1}$ then is represented by a polynomial $A(x)$ of degree $n(n+1)$. Now

$$A(x) = Q(x)(x-1)^2 + c(x-1) + d, \tag{8}$$

for a unique polynomial $Q(x)$ in $\mathbb{Z}[x]$, and unique elements $c, d \in \mathbb{Z}$. At this point the proof is straightforward: Compute c, d, and $Q(x)$ and show that c, d, and the coefficients of $Q(x)$ are nonnegative. The computations are not hard since division by $x - 1$ is simple: If $A(x) = B(x)(x-1) + d$, the coefficients of $B(x)$ are convenient sums of coefficients of $A(x)$, and $d = A(1)$. Then division of $B(x)$ by $x - 1$ completes the argument; the only nonroutine step is proving that the coefficients of the quotient are nonnegative.

Theorem 1.6 can be strengthened. Applying the theorem to the function $-f$ shows that if f' and f'' are nonpositive, then E_{n+1} is again at least as good an estimate of $\int_0^1 f(x)\,dx$ as E_n is. A proof similar to that of Theorem 1.6 shows that if f' is nonnegative and f'' is nonpositive the same conclusion holds. This is all summarized in the next theorem. In this case, subtract the second differences with negative coefficients, starting at the *smallest* argument rather than at the largest one.

THEOREM 1.7. *Let f be defined on* $[0,1]$ *and be twice differentiable. If* f' *and* f'' *do not change sign, then* $E_n \leq E_{n+1} \leq \int_0^1 f(x)\,dx$ *or* $E_n \geq E_{n+1} \geq \int_0^1 f(x)\,dx$ *for all n.*

After obtaining Theorem 1.7, I played a little with trapezoidal estimates and conjectured that if neither $f^{(2)}$ nor $f^{(3)}$ changes sign, then these estimates approach the integral monotonically. Moreover, an argument similar to the one for the Riemann sum estimates, but using third and second differences, looked like it would work but be quite messy. It also looked as if neither $f^{(4)}$ nor $f^{(5)}$ changes sign, then the Simpson estimates also approach the integral monotonically. A pattern emerged: If an approximation method was exact for polynomials of degree at most k then it would be monotonic if $f^{(k+1)}$ and $f^{(k+2)}$ don't change sign. Before attempting to prove such a broad conjecture, I thought it prudent to be sure it was not already a theorem, perhaps a century old but forgotten. In response to inquiries, T. Rivlin mentioned a paper of D. J. Newman [4] and one of his own that examine the monotonicity of the midpoint method [6], both published in 1974. Following this lead, I found that "monotonicity" questions had been investigated as far back as 1929. I will describe two elementary approaches in detail and state Newman's result, which confirmed my conjecture for a broad class of estimation techniques.

2. Reduction to Simpler Functions

The earliest appearance of the monotonicity problem in print was in 1950 in a paper of P. Turán [11]. In connection with the study of the zeros of the Legendre polynomials, he cited "some unpublished theorems of Fejér, which I mention here with his kind permission; he found his results during the winter term of the academic year 1928/1929." Theorem 2.1 below is one of the four results of Fejér, the others obtained by other combinations of "nonincreasing" or "nondecreasing" with "concave" or "concave down." In 1961 Turán and Szego published a proof in a joint paper [10]. Their approach depends on an observation that substantially reduces the class of functions to be examined.

Consider the class C of decreasing functions $f(x)$ on $[0,1]$ that are convex ("concave up") in the sense that they lie below their chords. For convenience assume $f(1)=0$.

This class includes the special functions

$$f_r(x) = \begin{cases} r-x & \text{for } 0 \le x \le r \\ 0 & \text{for } r \le x \le 1, \end{cases} \tag{9}$$

for all $0 \le r \le 1$.

Any function in C can be uniformly approximated (from below) by a suitable linear combination of the functions $f_r(x)$ with *positive* coefficients. FIGURE 1 illustrates why this is the case.

Starting at $(1,0)$ move to the left to $(r_1,0)$ until the curve is ε above the x-axis. From $(r_1,0)$ draw the tangent to the curve and continue it to the left until it is ε below the curve, at a point with x coordinate r_2. Continue constructing such segments of tangents until meeting the y-axis, obtaining a finite sequence of numbers $r_1, r_2, \dots, r_k$. Then there are positive constants $c_1, \dots, c_k$ such that $|f(x) - \sum_{i=1}^{k} c_i f_{r_i}(x)| \le \varepsilon$ for all x in $[0,1]$. This is the key to Theorem 2.1, which had been stated by Fejér.

Of the four classes of functions to consider, we take only the case $f(x)$ "decreasing" and "concave up."

The theorem concerns two point systems in $[0,1]$, namely

$$0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1$$

and

$$0 = y_0 < y_1 < \cdots < y_{n+1} < y_{n+2} = 1.$$

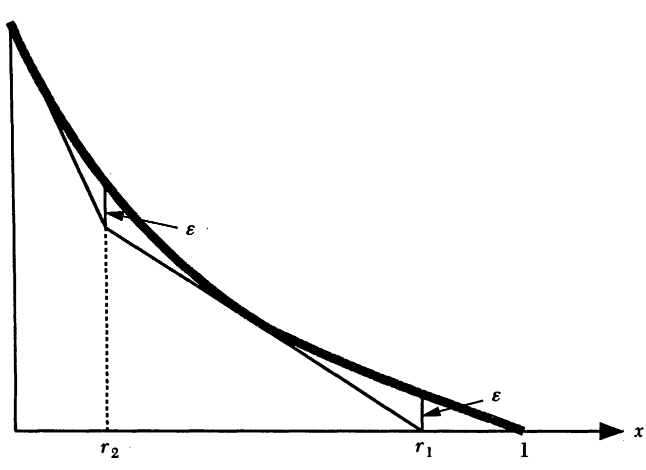

FIGURE 1

The first divides $[0,1]$ into $n+1$ sections and the second into $n+2$ sections. Moreover, we assume that the point systems interlace each other, that is,

$$x_i < y_{i+1} < x_{i+1},$$

$i = 0, 1, \ldots, n$, as shown in FIGURE 2.

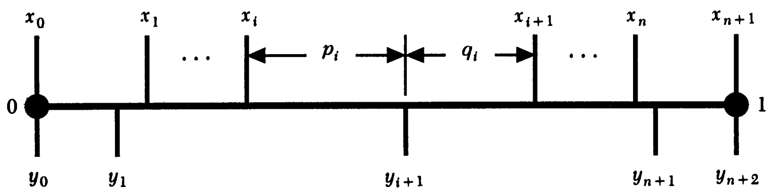

FIGURE 2

Let $p_i = y_{i+1} - x_i$ and $q_i = x_{i+1} - y_{i+1}$, $i = 0, 1, \ldots, n$, as shown in FIGURE 2. (If both the x_i's and the y_i's are equally spaced, they form an interlaced system. A simple computation shows that in this case $p_0 \geq p_1 \geq p_2 \geq \cdots \geq p_n$.) Let

$$S_n(f) = \sum_{i=0}^{n} f(x_i)(x_{i+1} - x_i) \quad \text{and} \quad S_{n+1}(f) = \sum_{i=0}^{n+1} f(y_i)(y_{i+1} - y_i),$$

the "left-hand" Riemann sums formed with these two subdivisions. We take this case since it is the one treated by Szegó and Turán. They prove the following theorem, which they credit to Fejér.

THEOREM 2.1. *In order that $S_{n+1}(f) \geq S_n(f)$ holds for a fixed n and all nonincreasing concave up functions, it is necessary and sufficient that*

$$\sum_{i=0}^{j} q_i(p_{i+1} - p_i) \leq 0 \tag{10}$$

for each j, $0 \leq j \leq n$.

Sketch of proof (Assume that $f(1) = 0$). Since any nonincreasing concave up function in $[0,1]$ can be uniformly approximated by the special functions (9), we restrict our attention to these functions. The idea is to show that the set of $n+1$ inequalities (10) is equivalent to the inequalities

$$S_{n+1}(f_r) \leq S_n(f_r) \tag{11}$$

for all r, $0 \leq r \leq 1$. To do this, first evaluate both sides of (11).

Given r, $0 \leq r \leq 1$, define j by

$$x_j \leq r < x_{j+1}.$$

If $r = 1$ define j to be $n+1$. Then define a by the equation

$$r = x_j + a.$$

Thus, if $j > 1$, we have

$$r = p_0 + q_0 + p_1 + q_1 + \cdots + p_{j-1} + q_{j-1} + a,$$

and, if $j = 0$, then $r = a$. Also we have

$$0 \leq a < p_j + q_j.$$

We first evaluate $S_n(f_r)$ with the aid of FIGURE 3.

If $j=0$, $S_n(f_r)=ax_1$. If $1\le j\le n$, then $S_n(f_r)$ can be viewed as the sum of the area of the large triangle of area $r^2/2$, some smaller triangles above it, and a trapezoid. We obtain

$$S_n(f_r)=\frac{r^2}{2}+\frac{1}{2}\sum_{i=0}^{j-1}(p_i+q_i)^2+\left(p_j+q_j-\frac{a}{2}\right)a,$$

which gives

$$S_n(f_r)=\frac{r^2-a^2}{2}+\frac{1}{2}\sum_{i=0}^{j-1}p_i^2+\sum_{i=0}^{j-1}p_iq_i+\frac{1}{2}\sum_{i=0}^{j-1}q_i^2+(p_j+q_j)a. \tag{12}$$

The evaluation of $S_{n+1}(f_r)$ is similar but splits into two cases, $a\le p_j$ and $a>p_j$. We leave it for the reader to treat these two cases and complete the proof.

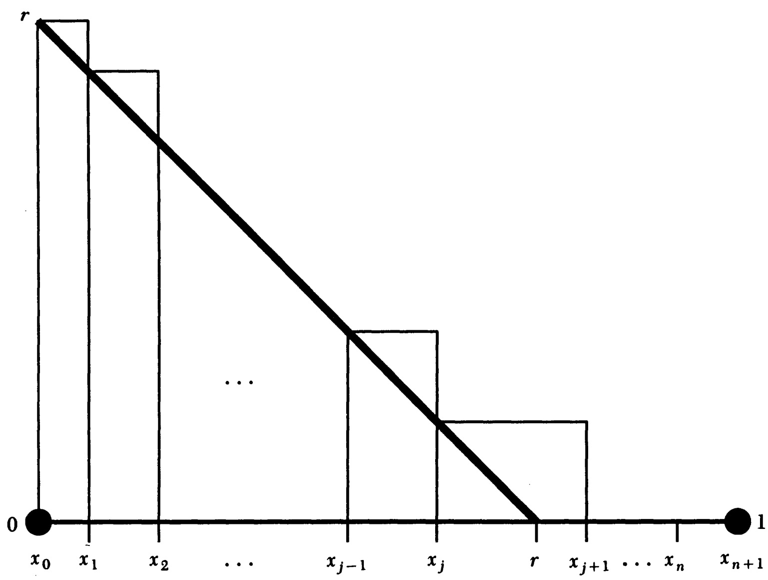

FIGURE 3

3. Jensen's Inequality

In response to my question, A. Pinkus sent me a different approach in May 1992, which exploits the fact that by definition, a convex function lies below its chords. Analytically, this condition reads as follows.

Let $a\le c\le b$. There are unique numbers p and q such that $0\le p,q\le 1$, $p+q=1$, and $c=pa+qb$. Then a function is convex on the interval $[a,b]$ if, and only if,

$$f(pa+qb)\le pf(a)+qf(b) \tag{13}$$

for all such p and q. (The special case when c is the midpoint of $[a,b]$, hence $p=1/2=q$, was used in Section 1.)

THEOREM 3.1. *Let the function f defined on* $[0,1]$ *be convex and nonincreasing. Then, in the notation of Theorem* 1.6, $E_n\le E_{n+1}$.

Sketch of proof. Note that $i/(n+1) < i/n < (i+1)/(n+1)$, and we have

$$\frac{i}{n} = \frac{n-i}{n} \cdot \frac{i}{n+1} + \frac{i}{n} \cdot \frac{i+1}{n+1}.$$

By (13)

$$f\left(\frac{i}{n}\right) \le \frac{n-i}{n} f\left(\frac{i}{n+1}\right) + \frac{i}{n} f\left(\frac{i+1}{n+1}\right). \tag{14}$$

Applying (14) to each summand in $\sum_{i=1}^{n}(1/n)f(i/n)$ will complete the proof.

4. The Generalization to All Equi-spaced Newton-Cotes Estimates

The right-hand, left-hand, midpoint, trapezoidal, and Simpson's methods are all special cases of *Newton-Cotes quadratures*. These quadratures estimate $\int_0^1 f(x)\,dx$ as follows.

Let k be a nonnegative integer and $x_0 < x_1 < \cdots < x_k$ be $k+1$ points in $[0,1]$. (In applications, usually $x_0 = 0$, $x_k = 1$ and the points are equally spaced.) Let $p(x)$ be the unique polynomial of degree at most k such that $p(x_i) = f(x_i)$, $0 \le i \le k$. Then the Newton-Cotes estimate of $\int_0^1 f(x)\,dx$ is $\int_0^1 p(x)\,dx$. Note that the estimate is exact when $f(x)$ is a polynomial of degree $\le k$. Equivalently, one chooses fixed "weights" $a_0, a_1, \dots, a_k$ such that the estimate $\sum_{i=0}^{k} a_i f(x_i)$ equals $\int_0^1 f(x)\,dx$ for $f(x) = x^0, x^1, \dots, x^k$.

For $k = 0$ we get the left-hand, right-hand, or midpoint estimates according as x_0 is chosen to be 0, 1/2, or 1, respectively. For $k = 1$, $x_0 = 0$, $x_1 = 1$ we get the trapezoidal estimate (with a single trapezoid) and for $k = 2$, $x_0 = 0$, $x_1 = 1/2$, $x_2 = 1$, Simpson's estimate (with a single parabola).

Consider a fixed Newton-Cotes method. Let n be a positive integer and break the interval $[a, b]$ into n sections of equal lengths. Estimate the integral of $f(x)$ over each of the n sections by a single application of that method. The sum of these n "local" estimates is called the *composite estimate* of $\int_a^b f(x)\,dx$, and is denoted $L(n)$. (The trapezoidal and Simpson's methods presented in freshman calculus are examples of composite estimates.)

In 1974 D. J. Newman [4] obtained the following general theorem.

THEOREM 4.1. *Let k be a positive integer and consider the Newton-Cotes procedure based on k sections of equal length. Let L_n denote the corresponding composite estimate of $\int_a^b f(x)\,dx$. Assume that $f^{(k+1)}$ and $f^{(k+2)}$ are continuous and do not change sign in $[a, b]$: Then the estimates $L_1, L_2, \dots, L_n, \dots$ approach $\int_a^b f(x)\,dx$ monotonically.*

For the proof, which depends on advanced calculus, we refer the reader to the paper itself.

5. Questions That Remain

In Newman's theorem the underlying Newton-Cotes method is based on $k+1$ equally spaced points $x_0, x_1, \dots, x_k$, with x_0 and x_k being the ends of the interval. It does not include the midpoint method, for example. Is Theorem 4.1 a special case of a more general theorem? Is it true that if a Newton-Cotes procedure is exact for polynomials of degree k but not of degree $k+1$ and neither $f^{(k+1)}$ nor $f^{(k+2)}$ changes sign, then the estimates L_n converge monotonically?

Might the following known generalization of Jensen's inequality, stated as Theorem 5.2, do the trick, along the lines of the proof of Theorem 3.1?

THEOREM 5.2. *Let $f(x)$ have a nonnegative nth derivative everywhere and let $a_0, a_1, \ldots, a_n$ be $n+1$ distinct real numbers. Then*

$$\sum_{i=0}^{n} \frac{f(a_i)}{\prod_{\substack{0 \le j \le n \\ j \ne i}} (a_i - a_j)} \ge 0.$$

The case $n = 2$ is Jensen's inequality.

Could the technique of reducing the general case to a special case of functions, as in the proof of Theorem 2.1, be generalized? For instance, is there a way of approximating functions for which $f^{(2)}$ and $f^{(3)}$ don't change sign, by a sum of second-degree polynomials?

What about the algebraic approach? It provides an algorithm for testing any specific case as long as the points of subdivision are rational. But lurking behind this algebraic equation are several concerning the group $\mathbb{Z}^n$. For instance, *when is an element in $\mathbb{Z}^n$ expressible as a linear combination of "kth differences" with nonnegative coefficients*? (An element in $\mathbb{Z}^n$ of the form $(0, \ldots, 0, 1, k, \ldots, k, 1, 0, \ldots, 0)$, where the nonzero entries are the numbers $(-1)^i \binom{k}{i}$, corresponds to a kth difference with difference 1.) Of interest in our problem is the question, when is an element in $\mathbb{Z}^n$ expressible as a linear combination of $(k+1)$st differences with nonnegative (nonpositive) coefficients and $(k+2)$nd differences with nonnegative (nonpositive) coefficients?

After all this, a calculus student might ask, "But how does a *single* application of the Newton-Cotes method for k sections compare with a *single* application for $k+1$ sections? For instance, how does the trapezoidal estimate compare with Simpson's method?"

This question was considered by H. Brass [2] in 1978. Let Q_m be the Newton-Cotes estimate of $\int_a^b f(x)\,dx$ based on the $m+1$ equally spaced points $x_0 = a, x_1, x_2, \ldots, x_m = b$. He proved that if $f^{(2k)}(x) \geqslant 0$, then $Q_{2k-1} \ge Q_{2k} \ge Q_{2k+1}$. Thus, if all even order derivatives are nonnegative then the various Newton-Cotes estimates approach $\int_a^b f(x)\,dx$ monotonically.

Of course, the student then asks, "But how do the composite estimates based on trapezoids compare with a single Simpson estimate?" If T_k denotes the estimate based on k trapezoids of equal width and S_k denotes Simpson's estimate based on k parabolas of equal width, how does T_k compare with S_1? Using the algebraic approach, one may show that if $f^{(2)}(x) \ge 0$, then $T_i \ge S_1$, $i = 1$, 2, and 3.

This same persistent student might ask, "In the trapezoidal method the weights are $1/2$ and $1/2$. In Simpson's method they are $1/6, 4/6, 1/6$. Are they always positive in any Newton-Cotes procedure with equally spaced points?" Intuition may suggest "yes", and it would be right through the first seven cases. However, some of the weights are negative in the case of eight sections (that is, nine equally spaced points). With nine sections, all weights are positive. If more sections are used, then approximately half the weights are negative. (See [12].) The cancellation effect of positive and negative terms increases the importance of roundoff errors and is one of the reasons that higher order Newton-Cotes procedures are impractical.

Not only do some of the weights become negative in high order Newton-Cotes estimates, but some become very large. Though a_0 and a_k are asymptotic to $1/(k \log k)$ and a_1 and a_{k-1} are asymptotic to $1/(\log k)^2$, all the rest of the weights become arbitrarily large in absolute values as $k \to \infty$. For instance, for even k, $k = 2r$, a_r is asymptotic to

$$\frac{(-1)^{r-1} 4^r}{\sqrt{\pi}\,(\log 2r)^2}.$$

(See **[13]**.)

Other questions may occur to the reader. (For the monotonicity of the Gauss or Romberg estimates see **[2]** and **[9]**. For the relation to convexity cones see **[1]** and **[14]**.)

In any case, when we assure our calculus students that "the estimates get better as we subdivide finer and finer," maybe we should add the caveat, "as a general trend."

Acknowledgment. I wish to thank Anthony Barcellos for preparing the illustrations (using Cohort software).

REFERENCES

1. D. Amir and Z. Ziegler, Convexity cones and their duals, *Pacific J. Math.* 27 (1968), 425–440.
2. H. Brass, Monotonie bei den Quadraturverfahren von Gauss and Newton-Cotes, *Numer. Math.* 30 (1978), 349–354.
3. D. Kincaid and W. Cheney, *Numerical Analysis*, Brooks/Cole Publishing Co., Belmont, CA, 1990, p. 478.
4. D. J. Newman, Monotonicity of quadrature approximations, *Proc. Amer. Math. Soc.*. 42 (1974), 251–257.
5. A. Pinkus, letter of May 11, 1992.
6. T. J. Rivlin, The Lebesgue constants for polynomial interpolation, in *Functional Analysis and Its Applications*, Ed. H. G. Garnir et al., Lecture Notes in Mathematics, 399 (1974), 422–437 (in particular, 428–429).
7. F. Stenger, Bounds on the error of Gauss-type quadratures, *Num. Math.* 8 (1966), 150–160.
8. J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag New York, Inc., New York (1979), 123–125.
9. T. Ström, Monotonicity in Romberg quadrature, *Math. Comp.* 26 (1972), 461–465.
10. G. Szegö and P. Turan, On the monotone convergence of certain Riemann sums, *Publ. Math. Debrecen* 8 (1961), 326–335.
11. P. Turán, On the zeros of the polynomials of Legendre, *Casopis Pest Mat. Fys.* 75 (1950), 113–122, (MR Vol 12), p. 824).
12. J. V. Uspensky, On the expansions of the remainder in the Newton-Cotes formula, *Trans. Amer. Math. Soc.* 17 (1935), 381–396.
13. ______, Sur les valeurs asymptotiques des coefficients de Cotes, *Bull. Amer. Math. Soc.* 31 (1925), 145–156.
14. Z. Ziegler, Generalized convexity cones, *Pacific J. Math.* 27 (1966), 561–580.