Partitions for Optimal Approximations

J. Rosenblatt

Department of Mathematics
University of Illinois
Champaign-Urbana, Illinois, USA
rosnbltt@illinois.edu

Abstract. The Riemann integral can be approximated using partitions and a rule for assigning weighted sums of the function at points determined by the partition. Approximation methods commonly used include endpoint rules, the midpoint rule, the trapezoid rule, Simpson’s rule, and other quadrature methods. The rate of approximation depends to a large degree on the rule being used and the smoothness of the function, but it also depends on the partition. We show that when one chooses an optimal partition, one gets a precise asymptotic rate of approximation and characteristic distribution of the points in the partition.

1. Introduction

There is an extensive literature on optimal approximation, extending from classical approximation theory to modern image processing. It would be difficult to give adequate references to cover this literature, so we give only those that we know are directly related to the results in this article. Despite all of this literature, it has been difficult to find articles about optimal approximation, and the distribution of the parameters that provide the optimal approximation, that use readily understandable, rigorous arguments. This may be due to this author’s lack of knowledge of the literature more than anything else. This particular article focuses on a special case of this type of problem: optimal partitions for approximation of the Riemann integral using given numerical rules.

First consider a very basic case of optimal approximation of the Riemann integral. Take a continuous function $f$ on $[0,1]$ and choose a partition $P_n =$
\{x_1, \ldots, x_n\} of [0, 1] containing \( n \) distinct points. We assume here that \( 0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1 \). Use the left endpoint rule \( E(f, 0, 1, P_n) = \sum_{k=0}^{n} f(x_k)(x_{k+1} - x_k) \) to obtain an approximation of \( \int_0^1 f(t) \, dt \). There is some, perhaps not uniquely determined, optimal partition \( P_n^\# \) which minimizes the error \( | \int_0^1 f(t) \, dt - E(f, 0, 1, P_n) | \) over all \( n \) point partitions \( P_n \). We would like to answer the following questions, in cases like this one and more generally:

1. How do we determine the optimal partition? When does it exist and when is it unique?
2. What is the overall error for the numerical approximation of the integral using the optimal partition as \( n \) tends to \( \infty \)?
3. How are the points of \( P_n^\# \) distributed as \( n \) goes to \( \infty \)?

Even if \( f \) is strictly increasing, the answers to these questions are not obvious. We can answer these questions at least when the function is also continuously differentiable.

To understand some of the issues here, let \( f \) be continuous and consider the lower Riemann sums \( L(f, 0, 1, P_n) = \sum_{k=0}^{n} \left( \min_{x \in [x_k, x_{k+1}]} f(x) \right)(x_{k+1} - x_k) \). If \( f \) were non-decreasing, this would of course be the left endpoint rule above. Assuming that the mesh of \( P_n \), that is the maximum value of \( x_{k+1} - x_k \), \( k = 0, \ldots, n \), tends to zero, we have \( L(f, 0, 1, P_n) \to \int_0^1 f(t) \, dt \) as \( n \to \infty \). If the mesh tends to zero slowly, then this convergence will be slow too. But if \( f \) is continuously differentiable, and one chooses the uniform partition \( P_n^U \), where the values \( x_{k+1} - x_k = \frac{1}{n+1} \) for all \( k \), then the error \( | \int_0^1 f(t) \, dt - L(f, 0, 1, P_n^U) | \) is bounded by \( \max_{x \in [0,1]} |f'(x)|/n \). One might think that one could get an even better rate of approximation by choosing the partition more specifically with the function in mind. However, Tasaki [7] computed the actual error, using the optimal partition \( P_n^\# \) for this rule, and showed that \( \lim_{n \to \infty} n \left| \int_0^1 f(t) \, dt - L(f, 0, 1, P_n^\#) \right| = \frac{1}{2} \left( \int_0^1 |f'(t)|^{1/2} \, dt \right)^2 \). In addition, Tasaki [7] obtains a result like this for the trapezoid rule. Actually, both of these results follow from the general method in McClure [6]. One of the goals of this article is to obtain asymptotic results of this type for general weighted rules for approximating the Riemann integral. We will use the ideas and results in McClure [6] to get this asymptotic result. In some cases the relevant result in [6] gives this asymptotic rate, but in general we need to apply the ideas in [6] rather than the theorems themselves.

We would like to determine explicitly the optimal partition \( P_n^\# \). But while we can write down recursive formulas and do some calculations in this direction, it seems to be difficult to determine this partition precisely in general. However, we can determine its distribution asymptotically as \( n \) tends to \( \infty \). It is intuitively clear that in regions where \( f \), or its derivatives \( f^{(k)} \), are changing quickly, one has to put more of the points of \( P_n^\# \) than in other regions, and
so the distribution of $P_n^#$ must relate in some fashion to the derivatives $f^{(k)}$. We propose that the best way to study the distribution of $P_n$ is to consider the probability measure $\nu(P_n) = \nu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}$ where $\delta_x$ denotes the point mass measure at $x$. The question is: does $\nu(P_n^#)$ converge weakly? That is, does the limit $\Lambda(h) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} h(x_k)$ exist for all continuous functions $h : [0, 1] \to \infty$? If this weak limit does exist, then $\Lambda(h) = \int_{[0,1]} h \, d\nu$ for some probability measure $\nu$ on $[0, 1]$. So the follow up question is: how is $\nu$ related to $f$? We can use McClure [6] directly to answer this question in some cases, for example in the cases considered by Tasaki [7]. But in general we will have to adapt the methods in [6] to get the results that we want. For example, using the lower Riemann integral estimates $L(f, 0, 1, P_n^#)$, when $f$ is continuously differentiable, the weak limit $\nu$ of $\nu(P_n^#)$ exists and is given by

$$\int_{[0,1]} h(t) \, d\nu(t) = \frac{1}{I} \left( \int_{0}^{1} h(t)|f'(t)|^{1/2} \, dt \right)$$

where $I$ is the normalizing factor $I = \int_{0}^{1} |f'(t)|^{1/2} \, dt$.

In Section 2, we set up and prove the basic estimates needed to get asymptotic rate results and to prove distributional limit theorems. In Section 3, we consider additional issues concerning the optimal partitions.

2. General rate arguments

Consider a general numerical rule for approximating the Riemann integral as follows. We take a discrete probability measure $\mu$ on $[0, 1]$ of the form $\mu = \sum_{s=1}^{S} c_s \delta_{\gamma_s}$ where $c_s > 0$ for all $s$, $\sum_{s=1}^{S} c_s = 1$, and $0 \leq \gamma_s \leq 1$ for all $s$. Then define the weight $W_\mu(f, a, b) = \left( \sum_{s=1}^{S} c_s f(a + \gamma_s(b - a)) \right)$ and the approximation $A_\mu = A_\mu(f, a, b)$ for $\int_{a}^{b} f(t) \, dt$ by

$$A_\mu(f, a, b) = W_\mu(f, a, b) (b - a) = \left( \sum_{s=1}^{S} c_s f((1 - \gamma_s)a + \gamma_s b) \right) (b - a).$$

Denote the error here by $e_\mu(f, a, b) = \int_{a}^{b} f(t) \, dt - A_\mu$. We will be adding these types of errors over many intervals, so we will actually need to use $|e_\mu(f, a, b)|$ to avoid unexpected cancelations of terms. Hence, it will also be useful to know what class of functions $f$ give us $\int_{a}^{b} f(t) \, dt \geq A_\mu(f, a, b)$ or $\int_{a}^{b} f(t) \, dt \leq A_\mu(f, a, b)$ for all $(a, b)$, in order to know that the signs of the errors are the same over all subintervals of the partition.

Here are some examples of the method above.

1. The left hand rule is where $A_\mu(f, a, b) = f(a)(b - a)$. So $\mu = \delta_0$. 

2. The midpoint rule is where \( A_\mu(f, a, b) = \int_a^b f(x) \, dx \). So \( \mu = \delta_1/2 \).
3. The trapezoid rule is where \( A_\mu(f, a, b) = \int_a^b f(x) \, dx \). So \( \mu = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1 \).
4. Simpson’s rule is where \( A_\mu(f, a, b) = \int_a^b f(x) \, dx \). So \( \mu = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1 \).

Methods like the lower Riemann sum \( L(f, a, b) = \min_{x \in [a, b]} f(x) \) are similar but not of the form \( A_\mu(f, a, b) \) for some \( \mu \). Generally, the anticipation is that the more terms in \( \mu \), the better chance one has of getting a good approximation of the integral.

Now consider the approximation for \( \int_a^b f(t) \, dt \) where we implement the approximations \( A_\mu \) over a partition \( P_n = \{ x_1, \ldots, x_n \} \) of \([a, b]\). Here we are taking \( a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b \). We will sometimes want to use only some of the intervals \( I_k = [x_k, x_{k+1}] \) arising from \( P_n \). So we introduce the notation \( I_n = \{ i_k : k = 0, \ldots, n \} \). Then the approximations \( e_\mu(f, x_k, x_{k+1}) = \int_{x_k}^{x_{k+1}} f(t) \, dt - A_\mu(f, x_k, x_{k+1}) \). We do not know necessarily whether \( e_\mu(f, x_k, x_{k+1}) \) is positive or negative, and this may vary over the interval of integration \([a, b]\). In any case, we denote the overall error bounds by \( E_\mu(f, a, b, I_n) = \sum_{k=0}^n e_\mu(f, i_k) = \sum_{k=0}^n e_\mu(f, x_k, x_{k+1}) \). Of course, the partition \( P_n \) and the set of partition intervals \( I_n \) depend on which points we are using, not just on the number of points. But in most case this will not cause confusion. When this might matter, we will introduce some additional notation.

In order to handle the signs of the local errors, we also consider \( E_\mu^+(f, a, b, I_n) = \sum_{k=0}^n e_\mu(f, i_k) \). Clearly, \( |E_\mu(f, a, b, I_n)| \leq E_\mu^+(f, a, b, I_n) \). If we have some knowledge of the functions and the measure \( \mu \), then we might be able to make this more precise. For example, suppose \( \mu = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1 \), giving the trapezoid rule.

If \( f \) is convex on \([a, b]\), then all of the terms \( A_\mu(f, x_k, x_{k+1}) \geq \int_{x_k}^{x_{k+1}} f(t) \, dt \), and \( E_\mu^+(f, a, b, I_n) = -E_\mu^+(f, a, b, I_n) \).

The first approximation fact that we need is a local one: how large is \( e_\mu(f, a, a + h) \) as \( h \to 0^+ \)? Depending on \( \mu \), we have a precise rule of the form \( A_\mu(f, a, b) = \int_a^b f(t) \, dt \) for all polynomials \( f \) of degree \( k < k_0 \) for some \( k_0 \geq 1 \) that depends only on \( \mu \). Of course, if \( f \) is constant, then we always have \( A_\mu(f, a, b) = \int_a^b f(t) \, dt \) because \( \sum_{s=1}^S c_s = 1 \). Note: to have consistency in our formulas, we take \( 0^0 \) to be defined to be \( 1 \). In order to see what the relevant formula will be, assume that \( f \) has a power series expansion \( \sum_{k=0}^\infty \frac{f^{(k)}(a)}{k!}(x-a)^k \) that holds on all of \([a, b]\). Applying the weighted sum \( \mu \) and integrating termwise
gives 
\[ e_\mu(f, a, b) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} \left( \frac{1}{k + 1} - \sum_{s=1}^{S} c_s \gamma_s^k \right) (b - a)^{k+1}. \]

This suggests that the relevant multiplier is \( C_\mu(k) = \frac{1}{k+1} - \sum_{s=1}^{S} c_s \gamma_s^k \). The value of \( k_0 \) is the smallest value such that \( C_\mu(k) = 0 \) for all \( k < k_0 \) and \( C_\mu(k_0) \) is not zero. That is, \( k_0 \) as the first value of \( k \geq 1 \) for which \( C_\mu(k) \neq 0 \). Note that implicitly here \( A_\mu(f, a, b) = \int_a^b f(t) \, dt \) for all homogeneous polynomials of degree \( k \) if and only if \( C_\mu(k) = 0 \). So in this generality, we would have 
\[ e_\mu(f, a, a + h) = \frac{f^{(k_0)}(a)}{k_0!} \left( \frac{1}{k_0 + 1} - \sum_{s=1}^{S} c_s \gamma_s^{k_0} \right) h^{k_0+1} + o(h^{k_0+1}) \text{ as } h \to 0^+. \]

**Remark 2.1.** For computational purposes, notice that if the \( \gamma_s \) are distinct, then the non-vanishing of the Vandermonde determinant tells us that we can find coefficients \( c_s \) which give us a measure \( \mu \) for which \( k_0 > S \).

This computation makes it easy to see what happens when we consider the more general case that \( f \in C^{(k_0)}([0, 1]) \), the functions with a continuous derivative up to order \( k_0 \). Using Taylor’s Theorem with the Lagrange form of the remainder, there is a continuous function \( c(x), a \leq c(x) \leq x \) such that
\[ f(x) = \sum_{k=0}^{k_0-1} \frac{f^{(k)}(a)}{k!} (x - a)^k + \frac{f^{(k_0)}(c(x))}{k_0!} (x - a)^{k_0}. \]
It follows that \( e_\mu(f, a, x) = \int_a^x \frac{f^{(k_0)}(c(t))}{k_0!} (t - a)^{k_0} \, dt - \sum_{s=1}^{S} c_s \gamma_s^{k_0} (x - a)^{k_0+1}. \) Then apply the assumption that \( f^{(k_0)} \) is continuous and the fact that \( c(x) \) is continuous, it is easy to see that one obtains the following:

**Proposition 2.2.** For \( f \in C^{(k_0)}([a, b]) \), and a probability measure \( \mu \) as above,
\[ e_\mu(f, a, a + h) = \frac{f^{(k_0)}(a)}{k_0!} C_\mu(k_0) h^{k_0+1} + o(h^{k_0+1}) \]
as \( h \to 0^+. \)

**Proof.** We can write 
\[ e_\mu(f, a, a + h) = \int_a^{a+h} f^{(k_0)}(c(t)) \, dt - \sum_{s=1}^{S} c_s \gamma_s^{k_0} h^{k_0+1} + \Delta(h) \]
So 
\[ e_\mu(f, a, a + h) = \int_a^{a+h} \frac{f^{(k_0)}(a)}{k_0!} (t - a)^{k_0} \, dt - \sum_{s=1}^{S} c_s \gamma_s^{k_0} h^{k_0+1} + \Delta(h) \]
where 
\[ \Delta(h) = \int_a^{a+h} \frac{f^{(k_0)}(c(t)) - f^{(k_0)}(a)}{k_0!} (t - a)^{k_0} \, dt - \sum_{s=1}^{S} c_s \gamma_s^{k_0} h^{k_0+1}. \]
Since \( f \in C^{(k_0)}([0, 1]) \) and \( c \) is continuous, as \( h \to 0^+ \), 
\[ \Delta(h) = o(1) \int_a^{a+h} (t - a)^{k_0} \, dt + o(1) \sum_{s=1}^{S} c_s \gamma_s^{k_0} h^{k_0+1} = o(h^{k_0+1}). \] But also clearly, 
\[ \int_a^{a+h} \frac{f^{(k_0)}(a)}{k_0!} (t - a)^{k_0} \, dt = \frac{f^{(k_0)}(a)}{k_0!} h^{k_0+1} + o(h^{k_0+1}). \]
For \( f \in C^{(k_0)}([a, b]) \) with \( f^{(k_0)} \) constant on \([a, b]\), and a probability measure \( \mu \) as above,

\[
e_{\mu}(f, a, a + h) = \frac{f^{(k_0)}(a)}{k_0} C_{\mu}(k_0) h^{k_0 + 1}\]

as \( h \to 0^+ \).

**Proof.** The error \( \Delta(h) = 0 \) is zero in this case when \( h > 0 \) is small enough. \( \square \)

Proposition 2.2 and Corollary 2.3 give this comparison result.

**Lemma 2.4.** Suppose \( f_1, f_2 \in C^{(k_0)}([a, b]) \) and \( |f_1^{(k_0)}| > |f_2^{(k_0)}| \) on \([a, b]\). Then if \( h > 0 \) is small enough, \( |e_{\mu}(f_1, a, a + h)| > |e_{\mu}(f_2, a, a + h)| \).

**Remark 2.5.** Proposition 2.2 shows that for \( f \in C^{(k_0)}([a, b]) \), the sign of \( e_{\mu}(f, a, a + h) \) for small values of \( h \) is determined by the sign of \( C_{\mu}(k_0) \). For example, with the midpoint rule, \( C_{\mu}(k) = \frac{1}{k+1} - \left( \frac{1}{2} \right)^k \). So \( k_0 = 2 \) and \( C_{\mu}(k_0) > 0 \). But for the trapezoid rule, \( C_{\mu}(k) = \frac{1}{k+1} - \frac{1}{2} \), so again \( k_0 = 2 \), but \( C_{\mu}(k_0) < 0 \).

If our function \( f \) has the sign of \( f^{(k_0)} \) the same on \([a, b]\), then \( |E_{\mu}(f, a, a, I_n)| = E_{\mu}^{+}(f, a, b, I_n) \) when the mesh of \( P_n \) is small enough. Sometimes though we do not even have to make the mesh small. For example, if \( f^{(2)} \) is always positive, then for the midpoint rule \( E_{\mu}(f, a, a, I_n) = E_{\mu}^{+}(f, a, a, I_n) \) while for the trapezoid rule \( E_{\mu}(f, a, a, I_n) = -E_{\mu}^{+}(f, a, a, I_n) \). With this in mind, we hope to estimate the asymptotics of \( E_{\mu}^{+}(f, a, b, I_n) \). Then we can restrict ourselves to the class of functions \( \Sigma(k_0) \subset C^{(k_0)}([a, b]) \) for which \( f^{(k_0)} \) does not change sign and we can use this estimate to give the estimate for \( |E_{\mu}(f, a, a, I_n)| \). This approach is technically necessary, but also has some real merit. For example, if \( f^{(k_0)} \) is both positive and negative on the interval, then we might find that \( \int_a^b f(t) \, dt = 0 \), and just because of sign cancelations we can compute the integral exactly using \( \sum_{k=0}^{n} A_{\mu}(f, x_k, x_{k+1}) \) for some small well placed partition.

This could also occur without the partition be uniquely determined.

We can use the local asymptotic estimates above, with additional assumptions about the partitions, to get an asymptotic estimate for \( E_{\mu}^{+}(f, a, b, I_n) \) of the form \( n^K E_{\mu}^{+}(f, a, b, I_n) \leq C(f) \) for some constant \( C(f) \). For example, this holds for the regular partitions as the mesh goes to zero. But in any case, at best we would only have the value \( K \) and an underestimate using \( C(f) \). Instead, we want to take an optimal partition \( P_n^\# \) and the associated intervals \( I_n^\# \), and then estimate \( E_{\mu}(f, a, b, I_n^\#) \) as \( n \) goes to \( \infty \). This will allow us to not only get the exact order \( K \) but also allow us to compute \( J(f) = \lim_{n \to \infty} n^K E_{\mu}^{+}(f, a, b, I_n^\#) \).

Here, by an optimal partition \( P_n^\# \), we mean a partition with \( n \) points such
that $E^+_\mu(f, a, b, I^*_n) \leq E^+_\mu(f, a, b, I_n)$ for all $n$ point partitions. In general, with an appropriate continuity assumption, the optimal partition at least exists. In the best cases, this partition is also uniquely determined. When existence of an optimal partition fails as it does sometimes below, we work with close to optimal partitions to compute $\inf_{\mathcal{P}_n} E^+_\mu(f, 0, 1, I_n)$, the infimum over the set $\mathcal{P}_n$ all $n$-point partitions:

In order to carry out our computation in general, we first consider a special class of functions for which the calculation is easier and which can be used to approximate our general function. Here is a simple example of what we mean. Consider a function $f$ which is a polynomial of degree $k_0$. Write $f(x) = Ax^{k_0} + q(x)$ where $q$ is a polynomial of degree less than $k_0$. Then by the definition of $C_{\mu}(k_0)$, $e_\mu(f, a, b) = \frac{A}{k_0!}C_{\mu}(k_0)(b - a)^{k_0+1}$. So $E_\mu(f, 0, 1, I_n) = \sum_{k=0}^{n} e_\mu(f, x_k, x_{k+1}) = \sum_{k=0}^{n} \frac{A}{k_0!}C_{\mu}(k_0)(x_{k+1} - x_k)^{k_0+1}$. Hence, $E^+_\mu(f, 0, 1, I_n) = |\frac{A}{k_0!}C_{\mu}(k_0)|\sum_{k=0}^{n} (x_{k+1} - x_k)^{k_0+1}$. This error is smallest, given the constraint

$\sum_{k=0}^{n} x_{k+1} - x_k = 1$, when the terms $x_{k+1} - x_k$ are all equal i.e. the partition $P^*_n$ in this case is uniformly distributed in $[0, 1]$.

Now consider a function $f$ such that $f^{(k_0)}$ is piecewise constant and right continuous. This is a class for which it is easier to carry out a computation but there is an issue with the optimal partition: it might not exist. We know that $\inf_{\mathcal{P}_n} E^+_\mu(f, 0, 1, I_n)$ exists, but it in general it might not be obtained. This is the situation for some of the choices of weights $\mu$ and functions $f$ that we are considering. We can work around this issue as follows. Suppose that $B_j$ are intervals partitioning $[0, 1]$ determined by $f$ and that $f^{(k_0)}$ is the constant value $A_j$ on $B_j$. These intervals are of the form $[\alpha_j, \beta_j)$ for $j = 1, \ldots, J$. We assume that $\alpha_1 = 0 < \beta_1 < \cdots < \alpha_J < \beta_J = 1$. There is no loss of generality in assuming that $A_j > 0$ for all $j$. Let $l_j$ be the length of $B_j$; so $\sum_{j=1}^{J} l_j = 1$.

We want to calculate $E^+_\mu(f, 0, 1, I_n)$ for some $P_n$ that gives a value close to $\inf_{Q_n \in \mathcal{P}_n} E^+_\mu(f, 0, 1, I(Q_n))$. First, suppose that $m_j$ points in $P_n$ are in $B_j$; so $\sum_{j=1}^{J} m_j = n$. The fact above about polynomials $f$ of degree $k_0$ tells us that to make $E^+_\mu(f, 0, 1, I_n)$ smaller, we should take the points in $B_j$ to be uniformly distributed. Indeed, if we were working in just one interval $B = [\alpha, \beta]$, then the optimal partition would be such that when including the endpoints $\alpha$ and $\beta$, then the points in the partition were uniformly distributed. So, taking into account that we can use 0 in the first interval and 1 in $B_j$ the actual best placement of the $m_j$ points using $[\alpha_j, \beta_j]$ would be to have one point being $\alpha_j$, one point being $\beta_j$ and the other $m_j - 2$ points evenly distributed between these.
In the actual situation because $B_j$ is only closed on the left, the best we can do initially is to place one point at $\alpha_j$ and one point close to $\beta_j$. The canonical choice we make is to place this point at $\alpha_j + (1 - \epsilon)(\beta_j - \alpha_j)$. Then we place the rest of the partition points uniformly between these two points. We refer to the resulting partition $P_n^*$ as an $\epsilon$-close to optimal partition for $f$.

We want to calculate the local error rates for each of the partition intervals. Rather, we can only use our local error rates for the partition intervals in $I_n(\epsilon)$, those that are not the ones of the form $I_j = [\alpha_j + (1 - \epsilon)(\beta_j - \alpha_j), \alpha_{j+1}]$. By a suitable choice of $\epsilon$, we can control the errors at the jump discontinuities so they contribute only $o(1/n^{k_0})$ at each value $j$. So we can arrange for the error from a jump discontinuity to be $O(\epsilon/n^{k_0})$. Since there are a finite number of jump discontinuities, using the local error rates gives

$$E_\mu^+(f, 0, 1, I_n) = E_\mu^+(f, 0, 1, I_n(\epsilon)) + O(\epsilon/n^{k_0})$$

By a suitable choice of $\epsilon$ gives a specific asymptotic rate as $\epsilon \to 0^+$.

Because $\sum_{j=1}^J m_j = n$, this is a constrained optimization problem with whole number variables. We can at least solve this constrained optimization asymptotically by first taking the values of $m_j$ to be real variables. A Lagrange multiplier calculation shows that the optimal values of $m_j = n \frac{A_j^{1/(k_0+1)}(1-\epsilon)l_j}{\sum_{i=1}^J A_i^{1/(k_0+1)(1-\epsilon)l_i}}$. This gives a specific asymptotic rate as $\epsilon \to 0^+$ and $n \to \infty$. Indeed, proceeding with these real values gives

$$\left| \frac{C_\mu(k_0)}{k_0!} \right| \sum_{j=1}^J A_j m_j (1 - \epsilon)l_j^{k_0+1} = \frac{1}{n^{k_0}} \left| \frac{C_\mu(k_0)}{k_0!} \right| \left( \sum_{j=1}^J A_j^{1/(k_0+1)}(1 - \epsilon)l_j \right)^{k_0+1}$$

$$= \frac{1}{n^{k_0}} \left| \frac{C_\mu(k_0)}{k_0!} \right| \left( \int_{0,1\setminus D_\epsilon} |f^{(k_0)}(t)|^{1/(k_0+1)} dt \right)^{k_0+1}$$

where $D_\epsilon$ denotes the union of the intervals $[\alpha_j + (1 - \epsilon)(\beta_j - \alpha_j), \alpha_{j+1}]$.

Now let $\epsilon \to 0^+$ and $n \to \infty$, we see that

$$n^{k_0} \left| \frac{C_\mu(k_0)}{k_0!} \right| \sum_{j=1}^J A_j m_j (1 - \epsilon)l_j^{k_0+1} \to \left| \frac{C_\mu(k_0)}{k_0!} \right| \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0+1)} dt \right)^{k_0+1}.$$
In this calculation, if one uses whole number values that are close to \(m_j\), then it is easy to see that the asymptotic result is the same. Indeed, choose whole numbers \(m_j^0\) and \(\delta_j\) such that \(m_j^0 + \delta_j = m_j\) and \(\sum_{j=1}^{J} m_j^0 = n\). This can be arranged with all \(\delta_j\) in \([-1, 1]\) except possibly for one value in \([-2, 2]\). Then we have

\[
\frac{|C_\mu(k_0)|}{k_0!} \sum_{j=1}^{J} A_j m_j^0 (\frac{(1 - \epsilon)l_j}{m_j^0})^{k_0 + 1} = 1 \frac{|C_\mu(k_0)|}{k_0!} \left( \sum_{j=1}^{J} A_j (\frac{(1 - \epsilon)l_j}{m_j^0})^{k_0 + 1} \right) \left( \sum_{j=1}^{J} A_j^{1/(k_0 + 1)} (1 - \epsilon)l_j \right)^{k_0}.
\]

Hence, using such values of \(m_j^0\), and letting both \(\epsilon \to 0^+\) and \(n \to \infty\), shows that

\[
\lim_{n \to \infty} n^{k_0} \inf_{P_n} E_\mu^+(f, 0, 1, I_n) = \frac{|C_\mu(k_0)|}{k_0!} \left( \sum_{j=1}^{J} A_j^{1/(k_0 + 1)} l_j \right)^{k_0 + 1}
\]

\[
= \frac{|C_\mu(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt \right)^{k_0 + 1}.
\]

This calculation gives us the following special case of Proposition 2.10.

**Lemma 2.6.** Assume that \(f \in C^{(k_0)}([0, 1])\) and that \(f^{(k_0)}\) is piecewise constant and right continuous. Then \(\lim_{n \to \infty} n^{k_0} \inf_{P_n} E_\mu^+(f, 0, 1, I_n) = \frac{|C_\mu(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt \right)^{k_0 + 1}\).

Also, if \(P_n^*\) is an \(\epsilon\)-close to optimal partition as above, then \(\nu(P_n^*)\) converges weakly as \(n \to \infty\) and \(\epsilon \to 0^+\). The limit measure \(\nu\) is such that for all continuous functions \(h : [0, 1] \to \infty\),

\[
\int_0^1 h(t) d\nu(t) = \frac{1}{I} \int_0^1 h(t) |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt
\]

where \(I\) is the normalizing constant \(\int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt\).

**Proof.** For the first limit, the argument is given above. For the weak limit comment, first observe that for the \(\epsilon\)-close to optimal partitions, the points in \(B_j\) are uniformly distributed. This fact and the values of \(m_j\) that solve the constrained optimization problem give the weak limit. \(\square\)

**Remark 2.7.** Suppose \(f^{(k_0)}\) is piecewise constant and right continuous. Fix \(\epsilon\) and for any arbitrary partition \(P_n\) of \([0, 1]\), let \(P_n(\epsilon)\) be the partition whose intervals do not intersect \(D_\epsilon\), the union of all the intervals of width \(2\epsilon\) around a jump discontinuity of \(f\). Let \(n_\epsilon\) be the number of partition points in \(P_n(\epsilon)\).
Let $I_\epsilon = \int_{[0,1]} |f^{(k_0)}(t)|^{1/(k_0+1)} dt$. It follows from an argument just like the one giving Lemma 2.6 that \( \lim_{n \to \infty} (n_\epsilon)^{k_0} \inf_{P_n} E^+_{\mu}(f, 0, 1, I_n(\epsilon)) \) exists and equals

$$\left( \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_{[0,1]} |f^{(k_0)}(t)|^{1/(k_0+1)} dt \right) \right)^{k_0+1} \cdot I(\epsilon).$$

In addition, as $n$ tends to $\infty$, $\nu(P_n^*(\epsilon))$ converges to the measure $\nu(\epsilon)$, supported on $[0,1] \setminus D_\epsilon$, such that for all continuous functions $h : [0,1] \to \infty$,

$$\int_0^1 h(t) d\nu(\epsilon)(t) = \frac{1}{I(\epsilon)} \int_{[0,1]} h(t) |f^{(k_0)}(t)|^{1/(k_0+1)} dt$$

where $I(\epsilon)$ is the normalizing factor $\int_{[0,1]} |f^{(k_0)}(t)|^{1/(k_0+1)} dt$. Now letting $\epsilon \to 0^+$ in these results actually gives Lemma 2.6 itself.

We now use this special case to prove a more general result. First, we prove an upper estimate for the overall error rate, and then we prove a lower estimate for the overall error rate.

**Lemma 2.8.** For all $f \in C^{(k_0)}([0,1])$,

$$\limsup_{n \to \infty} n^{k_0} E^+_{\mu}(f, 0, 1, I_n^\#) \leq \left( \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0+1)} dt \right) \right)^{k_0+1}.$$

**Proof.** We simplify this proof with no loss of generality by assuming $f^{(k_0)} > 0$ on $[0,1]$. Choose a right continuous, step function $g_2$ such that $f^{(k_0)} < g_2$ on $[0,1]$. By integrating, this gives a function $G_2$ in $C^{(k_0-1)}$ such that $G_2^{(k_0)} = g_2$ on $[0,1]$, except at the jump discontinuities of $g_2$.

Choose a partition $P_n^*$ which is $\eta$-close to giving $\inf_{P_n} E^+_{\mu}(G_2, 0, 1, I_n)$. By increasing $n$ if necessary, we can guarantee that the mesh of $P_n^*$ is small enough so that Proposition 2.4 tells us that $e(f, x_k, x_{k+1}) < e(G_2, x_k, x_{k+1})$ for all $x_k \in P_n^*$, except possibly for the $[x_k, x_{k+1}]$ on which $g_2$ has a jump discontinuity at $x_{k+1}$. Suppose $n$ is fixed and $[x_k, x_{k+1}]$ contains a jump discontinuity of $g_2$ at $x_{k+1}$. Since $f^{(k_0)}$ is continuous, by an appropriate choice of $\eta$, we can reposition $x_k$ closer to $x_{k+1}$ so that $e(f, x_k, x_{k+1}) \leq O(\epsilon/n^{k_0})$. Hence, for large enough $n$, with the appropriate choice of the positions of the partition point to the left of each jump discontinuity of $g_2$, we can arrange that $E^+_{\mu}(f, 0, 1, I_n^*) \leq E^+_{\mu}(G_2, 0, 1, I_n^*) + O(\epsilon/n^{k_0})$. Thus, for the optimal partition $P_n^\#$ for $f$, we have

$$n^{k_0} E^+_{\mu}(f, 0, 1, I_n^\#) \leq n^{k_0} E^+_{\mu}(G_2, 0, 1, I_n^*) + O(\epsilon).$$

Letting $n \to \infty$ and $\epsilon \to 0^+$, Lemma 2.6 gives

$$\limsup_{n \to \infty} n^{k_0} E^+_{\mu}(f, 0, 1, I_n^\#) \leq \left( \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_0^1 |G_2^{(k_0)}(t)|^{1/(k_0+1)} dt \right) \right)^{k_0+1}.$$. 
Now, we can also choose \( g_2 = G_2^{(k_0)} \) so that

\[
\left( \int_0^1 |G_2^{(k_0)}(t)|^{1/(k_0+1)} \, dt \right)^{k_0+1}
\]
is arbitrarily close to \( \left( \frac{|C_\mu(k_0)|}{k_0!} \int_0^1 |f^{(k_0)}(t)|^{1/(k_0+1)} \, dt \right)^{k_0+1} \). So it follows that

\[
\limsup_{n \to \infty} n^{k_0} E^+_\mu(f, 0, 1, I^n_\#) \leq \left( \frac{|C_\mu(k_0)|}{k_0!} \int_0^1 |f^{(k_0)}(t)|^{1/(k_0+1)} \, dt \right)^{k_0+1} \].

\[\square\]

**Lemma 2.9.** For all \( f \in C^{(k_0)}([0, 1]) \),

\[
\left( \frac{|C_\mu(k_0)|}{k_0!} \int_0^1 |f^{(k_0)}(t)|^{1/(k_0+1)} \, dt \right)^{k_0+1} \leq \liminf_{n \to \infty} n^{k_0} E^+_\mu(f, 0, 1, I^n_\#).
\]

**Proof.** We again simplify this proof with no loss of generality by assuming \( f^{(k_0)} > 0 \) on \([0, 1]\). Choose a right continuous, step function \( g_1, 0 < g_1 < f^{(k_0)} \) on \([0, 1]\). By integrating, this gives a function \( G_1 \) in \( C^{(k_0-1)} \) such that \( G_1^{(k_0)} = g_1 \) on \([0, 1]\) except at the jump discontinuities of \( g_1 \).

Take optimal partitions \( P^n_\# \) and associated sets of intervals \( I^n_\# \) that minimize \( E^+_\mu(f, 0, 1, I^n) \). Let \( X_D \) be the values of \( k \) such that \([x_k, x_{k+1}]\) contains a jump discontinuity of \( g_1 \). Let \( I_n \) be \([0, 1]\) with the intervals \([x_k, x_{k+1}]\), \( k \in X_D \), removed. Let \( n_D \) be the number of elements in \( X_D \). For large \( n \), with the mesh of \( P^n_\# \) small enough, each interval in \( X_D \) contains just one jump discontinuity of \( g_1 \). So \( n_D \) becomes the number of jump discontinuities of \( g_1 \) for large enough \( n \). Then for large enough \( n \), so that the mesh of \( P^n_\# \) is small enough,

\[
n^{k_0} E^+_\mu(f, 0, 1, I^n_\#) \geq n^{k_0} \sum_{k \notin X_D} e(G_1, x_k, x_{k+1}).
\]

On \( I_n \), which contains \( n - n_D + 1 \) intervals from \( I_n \), we can choose an optimal partition \( P^n_{n-n_D} \) and associated set of intervals \( I^n_{n-n_D} \) to minimize \( E^+_\mu(G_1 1_{I_n}, 0, 1, I^n_{n-n_D}) \) because the jump discontinuities of \( G^{(k_0)} \) are outside of \( I_n \). This is a restricted optimization as in Remark 2.7. Then we have

\[
n^{k_0} \sum_{k \notin X_D} e(G_1, x_k, x_{k+1}) \geq n^{k_0} E^+_\mu(G_1 1_{I_n}, 0, 1, I^n_{n-n_D}).
\]

So

\[
n^{k_0} E^+_\mu(f, 0, 1, I^n_\#) \geq n^{k_0} E^+_\mu(G_1 1_{I_n}, 0, 1, I^n_{n-n_D}).
\]

As \( n \to \infty \),

\[
n^{k_0} E^+_\mu(G_1 1_{I_n}, 0, 1, I^n_{n-n_D}) = \frac{n^{k_0}}{(n - n_D)^{k_0}} (n - n_D)^{k_0} E^+_\mu(G_1 1_{I_n}, 0, 1, I^n_{n-n_D})
\]
the optimal partitions for

Now, we can also choose \( g \) functions

Hence,

\[
\liminf_{n \to \infty} n^{k_0} E^+_{\mu_n}(f, 0, 1, I_n^\#) \geq \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt \right)^{k_0 + 1}.
\]

Proposition 2.10. For all \( f \in C^{(k_0)}([0, 1]) \), \( \lim_{n \to \infty} n^{k_0} E^+_{\mu_n}(f, 0, 1, I_n^\#) \) exists and

\[
\lim_{n \to \infty} n^{k_0} E^+_{\mu_n}(f, 0, 1, I_n^\#) = \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt \right)^{k_0 + 1}.
\]

If \( f \in \Sigma(k_0) \), then also \( \lim_{n \to \infty} n^{k_0} |E(f, 0, 1, I_n^\#)| \) exists and equals

\[
\lim_{n \to \infty} n^{k_0} |E(f, 0, 1, I_n^\#)| = \frac{|C_{\mu}(k_0)|}{k_0!} \left( \int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt \right)^{k_0 + 1}.
\]

The overall error rates can be used to give us the limiting distribution of optimal partitions. This result is actually derived just from Lemma 2.8 and Lemma 2.6, with some lower estimates that are reminiscent of the proof of Lemma 2.9.

Proposition 2.11. Suppose \( f \in C^{(k_0)}([0, 1]) \) and \( f^{(k_0)} \neq 0 \) on \([0, 1]\). Let \( P_n^\# \) be the optimal partitions for \( E^+_{\mu_n}(f, 0, 1, I_n) \). Then the probability measures \( \nu(P_n^\#) \) converge weakly and the limit is the measure \( \nu \) such that for all continuous functions \( h : [0, 1] \to \infty \),

\[
\int_0^1 h(t) d\nu(t) = \frac{1}{I} \int_0^1 h(t)|f^{(k_0)}(t)|^{1/(k_0 + 1)} dt
\]

where \( I \) is the normalizing constant

\[
\int_0^1 |f^{(k_0)}(t)|^{1/(k_0 + 1)} dt
\]
Proof. Assume without loss of generality that \( f(k_0) > 0 \) on \([0, 1]\). We use the same method as in Lemma 5 in McClure [6], with some modification. For a function \( G \) with \( G^{(k_0)} \) bounded and Riemann integrable, let \( D_G(r) = \int_0^r |G^{(k_0)}(t)|^{1/(k_0 + 1)} \, dt \). As in Lemma 5 in [6], we use compactness to assume without loss of generality that \( \nu(P^\#) \) converges weakly to a measure \( \nu \) with distribution function \( F(t) = \nu([0, t]) \). Choose a sequence of functions \( (G_m) \) supported on \([0, 1]\), with \( 0 < G^{(k_0)}_m < |f^{(k_0)}| \) on \([0, 1]\), so that \( G^{(k_0)}_m \) is a right continuous step function, and such that \( \lim_{n \to \infty} D_{G_m}(1) = D_{|f|}(1) \). Take the sequence of optimal partitions \( P^\#_n \) for \( E^+_\mu(f, 0, 1, I_n) \). There is a dense set of values \( r \in [0, 1] \) which are not discontinuities of any \( G^{(k_0)}_m \) and are not partition points in any \( P^\#_n \). Fix such a value of \( r \) and take the partition point \( x_{kn} \in P^\#_n \) such that \( x_{kn} < r < x_{kn+1} \).

Fix \( m \) for now. For large enough values of \( n \), the partition interval containing \( r \) will be contained in a set where \( G^{(k_0)}_m \) is constant, and so the formulas for \( e_\mu(G_m, a, b) \) in Proposition 2.2 show that we have \( |e_\mu(G_m, x_{kn}, x_{kn+1})| \geq |e_\mu(G_m, x_{kn}, r) + e_\mu(G_m, r, x_{kn+1})| \) because \((x_{kn+1} - x_{kn})^{k_0+1} \geq (x_{kn+1} - r)^{k_0+1} + (r - x_{kn})^{k_0+1} \). Let \( X_D = X_D(m) \) be the values of \( k \) such that \([x_k, x_{k+1}] \) contains a jump discontinuity of \( G^{(k)}_m \). For sufficiently large values of \( n \), the mesh of \( P^\#_n \) is small enough so that we have

\[
n^{k_0} E^+_\mu(f, 0, 1, I^\#_n) \geq (n/k_n)^{k_0} (n - k_n)^{k_0} \left( \sum_{1 \leq k \leq k_n - 1, k \notin X_D} |e_\mu(G_m, x_k, x_{k+1})| + |e_\mu(G_m, x_{kn}, r)| \right) + (n/(n - k_n))^{k_0} (n - k_n)^{k_0} \left( |e_\mu(G_m, r, x_{kn+1})| + \sum_{k_n+1 \leq k \leq n, k \notin X_D} |e_\mu(G_m, x_k, x_{k+1})| \right).
\]

Let \( I_n \) be \([0, 1] \) with the intervals \([x_k, x_{k+1}] \), \( k \in X_D \), removed. Take optimal partitions \( P(1)^*_n \) associated with the restricted minimization \( E^+_\mu(G_m 1_{I_n}, 0, r, I^*_n) \) and \( P(2)^*_{n-k_n} \) associated with the restricted minimization \( E^+_\mu(G_m 1_{I_n}, r, 1, I^*_n) \). We would have also

\[
n^{k_0} E^+_\mu(f, 0, 1, I^\#_n) \geq (n/k_n)^{k_0} (n - k_n)^{k_0} E^+_\mu(G_m 1_{I_n}, 0, r, I(1)^*_n) + (n/(n - k_n))^{k_0} (n - k_n)^{k_0} E^+_\mu(G_m 1_{I_n}, r, 1, I(2)^*_n).
\]

Letting \( n \to \infty \) forces both \( k_n \) and \( n - k_n \) to go to \( \infty \) too. By Lemma 2.8, \( \frac{|C^{(k_0)}|}{k_0} D_{|f|}(1)^{k_0+1} \geq \lim_{n \to \infty} \sup n^{k_0} E^+_\mu(f, 0, 1, I^\#_n) \). Also, letting \( n \to \infty \) and \( \epsilon \to 0^+ \), Lemma 2.6 gives

\[
\frac{|C^{(k_0)}|}{k_0} D_{G_m}(r)^{k_0+1} = \lim_{n \to \infty} (n/k_n)^{k_0} (n - k_n)^{k_0} E^+_\mu(G_m 1_{I_n}, 0, r, I(1)^*_n)
\]
and
\[
\frac{|C_\mu(k_0)|}{k_0!} (D_{G_m}(1) - D_{G_m(r)})^{k_0+1} = \lim_{n \to \infty} (n - k_n)^{k_0} E_\mu^+(G_m 1_{n_k}, r, 1, I(2)^*_{n_k}).
\]

At the same time, \(F(r) = \lim_{n \to \infty} k_n/n\) and \(1 - F(r) = \lim_{n \to \infty} (n - k_n)/n\). We know that \(F(r) > 0\) for large \(n\) because \(f^{(k_0)} > 0\) on \([0, 1]\). Hence, taking the limit as \(n \to \infty\) and \(\epsilon \to 0^+\), we have
\[
D_f(1)^{k_0+1} \geq F(r)^{-k_0} D_{G_m}(r)^{k_0+1} + (1 - F(r))^{-k_0} (D_{G_m}(1) - D_{G_m(r)})^{k_0+1}.
\]

Now let \(m \to \infty\). Then our inequality for the distribution functions becomes
\[
D_f(1)^{k_0+1} \geq F(r)^{-k_0} D_f(r)^{k_0+1} + (1 - F(r))^{-k_0} (D_f(1) - D_f(r))^{k_0+1}.
\]

Consider the function \(h(\alpha, \beta) = \alpha^{-k_0} \beta^{k_0+1} + (1 - \alpha)^{-k_0} (1 - \beta)^{k_0+1}\) where \((\alpha, \beta) \in (0, 1)^2\). This continuous function has a strict minimum of 1 at every point on the diagonal \(\alpha = \beta\) and is strictly larger than 1 off that diagonal. Hence, \(F(r) = D_f(r)/D_f(1)\). Since the values of \(r\) are dense, this proves our result.

**Remark 2.12.** In special cases, this result is in McClure [6]. We have adapted his argument to give the proof above in cases where we do not have the necessary subadditivity for \(e_\mu(f, a, b)\). We should observe that Proposition 2.11 in the special case of \(f^{(k_0)}\) being a step function was already proved in Lemma 2.6

**Remark 2.13.** It is well-known that any probability measure on \([0, 1]\) can be weakly approximated by a discrete measure. Indeed, it is possible to show that for any probability measure \(\nu\), there exists a sequence of partitions \(P_n\) whose mesh tends to zero as \(n\) tends to \(\infty\), such that \(\nu(P_n)\) converges weakly to \(\nu\) as \(n\) tends to \(\infty\). So, take partitions \(P_n\) and consider any numerical approximation \(\sum_{k=0}^n \sum_{s=1}^S c_s f(x_k + \gamma_s(x_{k+1} - x_k))(x_{k+1} - x_k)\) of \(\int_0^1 f(t) \, dt\), like the ones we have considered here. With the mesh of \(P_n\) tending to zero, this numerical approximation will converge to the Riemann integral for all continuous functions \(f\). But at the same time, we can have the measures \(\nu(P_n)\) converging weakly to whatever measure we would like (or even not converging at all). This makes it then even clearer how the distributional result in Proposition 2.11 depends critically on choosing the optimal partitions \(P_n\).

**Remark 2.14.** (a) McClure [6] proves this result under the additional assumption of subadditivity of the local error rates. For some classes of functions, and some weights \(\mu\), this holds. For example, if we apply the technique from McClure [6], we can get some asymptotic results that were proved later. The main result in Gleason [4] follows from McClure’s results once one knows a local asymptotic rate such as in Lemma 4 in [4]. Similarly, Theorem 1.2 and Theorem 1.4 in Tasaki [7] follow from McClure’s results and standard local
estimates for the lower Riemann sums and the trapezoid rule. See also Bronstein [1] for an extensive review of the literature on approximation of convex sets by polytopes.

(b) McClure’s results do not apply to Simpson’s rule quite as readily because it is not clear what class to use to guarantee subadditivity of $e_\mu(f, a, b)$.

Proposition 2.10 shows that if $f^{(4)} \geq 0$, then we do have a limiting value for $n^4 E_\mu(f, 0, 1, P_n^\#)$. It is $\frac{1}{2880} \left( \int_0^1 |f^{(4)}(t)|^{1/5} \, dt \right)^5$.

Remark 2.15. A more general version of the results above could be considered for $f$ which is just assumed to be strictly increasing. Then it is not clear what else is needed for an optimal partition to exist, when it is unique, or how to characterize the optimal partitions with a fixed number of points. It is not clear what type of local and overall error rates would hold, and if there would be distributional convergence for the optimal partitions.

3. Additional Issues

3.1. Transforming optimal partitions to uniform partitions. We have indicated that it is difficult to explicitly compute the optimal partitions even in fairly simple cases. But we have shown that we can compute the asymptotic distribution of these points. We can also see from this what transform needs to be applied to convert the optimal partitions into ones that are asymptotically uniformly distributed. We illustrate these issues here with the left-hand rule in general, and for some specific functions.

First, suppose $f$ is differentiable on $\mathbb{R}$. Assume that $f' > 0$ and so $f$ is strictly increasing. Take $P_n^\#$ to be a partition of $[0, 1]$ which maximizes the left endpoint Riemann sum $L(P_n) = \sum_{k=0}^n f(x_k)(x_{k+1} - x_k)$ i.e. minimizes the error in approximating the Riemann integral by the lower Riemann sum. The optimal partition must have the total differential $DL(f) = 0$ i.e. $\frac{dt}{dx_i} = 0$ for all $i = 1, \ldots, n$. This means that we have the following equations, for $i = 1, \ldots, n$:

$$f(x_{i-1}) + f'(x_i)(x_{i+1} - x_i) - f(x_i) = 0.$$

This formula does not immediately determine $x_1$; that value is only determined in the end using the fact that $x_{n+1} = 1$.

Remark 3.1. In particular, say $f$ in linear, given by $f(x) = bx$ where $b > 0$. Then

$$bx_{i-1} + b(x_{i+1} - x_i) - bx_i = 0$$

for all $i = 1, \ldots, n$. So $x_{i+1} - x_i = x_i - x_{i-1}$ for all $i = 1, \ldots, n$. This means that the points $x_i$ are uniformly distributed in $[0, 1]$. So in this case the asymptotic distribution of $P_n$ is the usual Lebesgue measure on $[0, 1]$. This is certainly what we expected from our earlier computations.
Remark 3.2. Moreover, in some cases, instead of having to optimize the choice of the partition points in aggregate, which is generally a computationally slow process, one can use the recursive formula to get a fairly efficient algorithm. Consider the case of powers \( f(x) = x^d \). Take an optimal partition \( P_n^\# = \{x_1(n), \ldots, x_n(n)\} \) with \( n \) points. Then our recurrence formula determining \( P_n^\# \) is

\[
x_{i+1}(n) - x_i(n) = \frac{x_i^d(n) - x_{i-1}^d(n)}{dx_{i-1}^d(n)}.
\]

For \( i, 2 \leq i \leq n \), in terms of the ratios \( r_i(n) \) defined by \( r_i(n) = x_{i+1}(n)/x_i(n) \), this formula becomes

\[
\frac{1}{d} = \frac{1}{dx_i(n)} \left( x_i(n) - \frac{x_{i-1}(n)}{r_{i-1}^d(n)} \right) = \frac{1}{d} \left( 1 - \frac{1}{r_{i-1}^d(n)} \right).
\]

That is,

\[
r_i(n) = \frac{d + 1}{d} - \frac{1}{dr_{i-1}^d(n)}.
\]

We can extend this formula to \( i, 1 \leq i \leq n \) by letting \( r_0(n) = (d+1)/d \). This is consistent with the recurrence formula for \( r_i(n) \) because we can formally take \( r_0(n) = x_1(n)/0 = \infty \) and \( (1/dr_0^d(n)) = 0 \). Now notice that we have for \( i = 1, \ldots, n \),

\[
x_i(n) = \frac{1}{\prod_{j=i}^n r_j(n)}.
\]

The advantage of using this formula to compute the points \( x_i(n) \) in the optimal partition \( P_n^\# \) is that when computing the points in \( P_{n+1}^\# \), the formulas give \( r_i(n+1) = r_i(n) \) for \( i = 1, \ldots, n \) and one need only compute the additional value

\[
r_{n+1}(n+1) = \frac{d + 1}{d} - \frac{1}{dr_{n}^d(n+1)}.
\]

Also, the new partition points \( x_i(n+1) \) are \( x_i(n)/r_{n+1}(n+1) \) for \( i = 1, \ldots, n \), and the one additional point \( x_{n+1}(n+1) \) is \( 1/r_{n+1}(n+1) \). These recursion formulas can be easily implemented to give quickly all the partition points for large values of \( n \).

Because we know the asymptotic distribution of \( \nu(P_n^\#) \), we can actually see what transform to use that would convert \( P_n^\# \) into an asymptotically uniform partition. An interesting aspect of this transform is that it does not convert \( P_n^\# \) to a precisely uniform partition. The transform that we need is a function \( x = \psi(y) \) that can be derived as follows. We know that for all continuous functions \( h \) on \([0, 1]\),

\[
\frac{1}{n} \sum_{k=1}^n h(x_k) \rightarrow \frac{1}{T} \int_0^T h(t) \sqrt{f(t)} \, dt \text{ as } n \rightarrow \infty.
\]

Here again
I is the normalizing constant $\int_0^1 \sqrt{f'(t)} \, dt$. So letting $t = \psi(s)$, use the change of variables formula for integrals to write

$$\frac{1}{T} \int_0^1 h(t) \sqrt{f'(t)} \, dt = \frac{1}{T} \int_0^1 h(\psi(s)) \sqrt{f'(\psi(s))} \psi'(s) \, ds$$

Thus, to transform the $x_k$ into something close to uniform distribution, we would want to have $\frac{d}{dt}(F \circ \psi)(s) = \frac{1}{T} \sqrt{f'(\psi(s))} \psi'(s) = 1$ for all $s$ where $F(t)$ is any antiderivative of $\frac{1}{T} \sqrt{f'(t)}$. This is the same as saying $F \circ \psi = s$ for all $s$ and so $\psi = F^{-1}$.

**Remark 3.3.** Finding $F$ explicitly could be difficult in general. But if $f(x) = x^d$, then you can easily compute that $F(t) = t^{(d+2)/2}$. Hence, $\psi(y) = y^{2/(d+2)}$.

So we are expecting that the partition points $x^\#_k$ for the optimal partition $P^\#_n$ for $f(x) = x^d$ will be very close to $\psi(k/(n+1) = (k/(n+1))^{2/(d+2)}$. For example, if $f(x) = x^d$, then we expect $x^\#_k$ to be approximately $(k/(n+1))^{1/2}$. One can easily check that this is not the precise value, but it is close. It would be worthwhile to know what the value of the error is. There are a number of ways to measure this, but here is one that may be interesting and worthwhile.

What is the asymptotic value of $\sum_{k=1}^n (x^\#_k - (k/(n+1))^2)$?

### 3.2. Nearly optimal partitions by equal factors.

Since we do not know the formula for the optimal partitions, we are restricted in using them for actual computations. There is an approach that circumvents this problem in McClure [6]. The same idea also appears in both Tasaki [7] and Gleason [4].

Here is how this works. Since we do not know the formula for the optimal partitions, we are restricted in using them for actual computations. There is an approach that circumvents this problem in McClure [6]. The same idea also appears in both Tasaki [7] and Gleason [4].

Using Gleason’s lemma, Tasaki [7] took a partition $P_n^\#$ such that the values $\sqrt{f'(x_k)}(x_{k+1} - x_k)$ are constant for $k = 1, \ldots, n$. This partition gives local error and global error rates that are asymptotically identical to the optimal partitions. See also Lemma 7 in McClure [6] for this idea in more general terms. But the computation of such balanced partitions may not be much easier than that for optimal partitions.

### 3.3. Monte-Carlo methods using the distribution of the optimal partitions.

An alternate approach would be the following. We illustrate this again in the case that we have the left endpoint rule in mind. Let $X_k(\omega), k = 1, \ldots, n$ be IID randomly variables, with values in $[0, 1]$, that are defined on a probability space $(\Omega, p)$. Assume that their distribution function $G(r) = p\{X_k \in [0, r]\}$ is given by $G(r) = \frac{1}{T} \int_0^r \sqrt{f'(t)} \, dt$. We then compute the approximation to $\int_0^1 f(t) \, dt$ by taking $\sum_{k=1}^n f(X_k(\omega))(X_{k+1}(\omega) - X_k(\omega))$.

**Question:** Does $n \left| \int_0^1 f(t) \, dt - \sum_{k=1}^n f(X_k(\omega))(X_{k+1}(\omega) - X_k(\omega)) \right|$ converge almost surely in $\omega$ to $\left( \int_0^1 f'(t)^{1/2} \, dt \right)^2$?
Acknowledgements: I would like to thank Vilmos Totik for getting me heading in the right direction on this problem. I thank Horoyuki Tasaki and Peter Gruber for helpful comments. I would also like to thank Donald McClure whose work has clearly been indispensable in obtaining the results in this paper. Lastly, the Illinois Geometry Lab team that worked with me on optimal partitions in Spring, 2012 helped by keeping my hopes up in trying to understand the nature and structure of optimal partitions. This team consisted of Yitian Chen, Yiwang Chen, Ilkyoo Choi, and Michael Lowney. They are the ones who invented the technique in Remark 3.2.

References


Received: June 1, 2013