Bernstein Polynomials:

We saw that Taylor expansion gives us a polynomial with an error term, of course, only when the function is differentiable so many times as required there. This led us to see if, in general, functions defined on an interval around zero have power series expansion. such a hope failed because we found examples of functions which is zero at $x = 0$ and all its derivatives exist and are zero at this point. The only power series then is the zero function; but we had a non-zero function. Thus such a function, as in that example, can not be explained in terms of power series.

The question that naturally arises is, even if we can not *exactly* express it as an infinte power series, are there polynomials close to it. This question gains importance because for the above function, Taylor polynomials fail to give anything. In fact the Taylor expansion of the function (of any order) around zero will always give you $f(x) = f(x)$, simply because all derivatives are zero when $x = 0$ and thus only the 'error' term remains which then must equal $f(x)$. Yes, given any continuous function f on a closed bounded interval there are polynomials which are as close to f as you want.

We know the following. If a is a real number and $\epsilon > 0$, then there is a rational number r in the interval $(a - \epsilon, a + \epsilon)$. We shall now prove a similar theorem about continuous functions. if f is a continuous function on a closed bounded interval and $\epsilon > 0$ then there is a polynomial P whose graph lies in the band $(f - \epsilon, f + \epsilon)$. You may recall that band means the set of points $\{(x, y) : f(x) - \epsilon < y < f(x) + \epsilon)$. This is the region in the (x, y)-plane obtained by taking parallel graphs at distance ϵ above and below the graph of f. We came across the band while discussing uniform convergence of sequences of functions.

In other words, the role of real number is played by continuous function; role of interval is played by the class of functions whose graphs lie in the ϵ-band around graph of f as described above; the role of rational is played by usual high school polynomial.

To put it analytically, we can find a polynomial P such that $|f(x) - P(x)| < \epsilon$ for every x in the closed bounded interval. This is same as saying that there is a sequence of polynomials P_n which converge uniformly on the interval to f. Of course polynomials are defined on all of R, but we are not saying about the sequence of polynomials outside the interval $[0, 1]$. They may converge or may not converge.

This is very satisfying because it says that general continuous function is not too far from a polynomial. However, you should bear in mind, this does not mean any thing in terms of differentiability properties. A polynomial is differentiable at every point, whereas a continuous function need not be differentiable at any point what-so-ever. As a consequence, we can not even ask if the polynomials are related to the Taylor polynomials simply because the function, we started with, need not be differentiable.

This theorem is due to Weierstrass (as is the concept of uniform convergence of sequence of functions). But the proof we give is due to Bernstein. Usually, this theorem is not done in a first course of Calculus. The reason why I am doing is the following. first, it is satisfying and has a clear intuitive meaning devoid of complicated maths. secondly, it reassures us that the continuous functions we have dveloped are really not too far from polynomials which you have learnt in high school (as long as you are thinking of a closed bounded interval). Also the theorem has a proof using only high school algebra, again devoid of any difficult maths. And it is important from analysis point of view. It also naturally fits in with the questions raised in connection with Taylor expansion.

Theorem: Let f be a real valued continuous function on $[0, 1]$ and $\epsilon > 0$. Then there is a polynomial P such that $|f(x) - P(x)| < \epsilon$ for every $x \in [0, 1]$.

First we observe an important property of continuous functions defined on $[0, 1]$.

Fact: f is a continuous function on $[0, 1]$. Given $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$ and $x, y \in [0, 1]$.

When we say that a function is continuus, we just mean that it is continuous at every point a. In turn, when we say that it is continuous at a, we mean that given $\epsilon > 0$ there is a $\delta > 0$ (which could depend not only on ϵ but also on the point a), so that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$. What we have said above is different. As soon as $\epsilon > 0$ is given, there is a $\delta > 0$ that works for every point a; in the sense, now if you take any point a and any x with $|x - a| < \delta$ then you will have $|f(x) - f(a)| < \epsilon$. Because of this, this property of continuous function is called *uniform continuity.*

Just to bring home the point, let us consider the function $f(x) = x^2$ on all of R. Let us try to verify continuity with $\epsilon = 1$. if our point is $a = 0$, then you can take $\delta = 1$ (you must verify this statement). if our point is $a = 10$ then $\delta = 1$ will not do, you must choose smaller δ, say $\delta = 1/21$. If our point is $a = 100$ then this δ also will not do, you need to choose a much much smaller δ. It is not difficult to see that as the point a gets larger and larger, you need to choose smaller and smaller δ which approaches zero as a gets larger and larger. in other words, you can not choose one $\delta > 0$ that works for all a. Remember $\epsilon = 1$ is fixed for all this discussion. what the fact above says is that such a thing can not happen if you had, instead of R, a closed bounded interval.

The proof of the fact is simple. If possible, fix an $\epsilon > 0$ for which we can not find a $\delta > 0$. Thus $\delta = 1/2$ or $1/2^2$ or $1/2^3$ etc will not serve our requirement. thus for each $n = 1, 2, 3 \cdots$ we ca find two points x_n, y_n so that $|x_n - y_n| < 1/2^n$ yet $|f(x_n) - f(y_n)| \geq \epsilon$. The sequence (x_n) being bounded, there is a subsequence $\{x_{n_1}, x_{n_2}, x_{n_3}, \cdots\}$ which converges to a point a, but then $y_i = (y_i - x_i) + x_i$ tells us that $\{y_{n_1}, y_{n_2}, y_{n_3}, \cdots\}$ also converges to a, but then the sequence $f(x_{n_1}) - f(y_{n_1}), f(x_{n_2}) - f(y_{n_2}), f(x_{n_3}) - f(y_{n_1}), \cdots$ converges to $f(a) - f(a) = 0$ whereas each term of the sequence is larger than ϵ in modulus. this is a contradiction. This completes the proof.

to proceed to the proof of the polynomial approximation theorem, we make a few observations. Through out below, just for these calculations, we take $x^0 = 1$ when $x = 0$ — just for now.

$$\sum_0^n \binom{n}{k} x^k (1-x)^{n-k} = 1; \qquad n \geq 1. \qquad (\spadesuit)$$

This is simply the binomial expansion for $(x + 1 - x)^n$.

$$\sum_0^n k \binom{n}{k} x^k (1-x)^{n-k} = nx; \qquad n \geq 1.$$

You can sum from $k = 1$ onward; then you can write $k\binom{n}{k} = n\binom{n-1}{k-1}$; then take nx outside the sum. recognize binomial expansion of $(x + 1 - x)^{n-1}$.

$$\sum_0^n k(k-1) \binom{n}{k} x^k (1-x)^{n-k} = n(n-1)x^2; \qquad n \geq 1.$$

Again, it suffices to sum only for $k \geq 2$; write $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$ etc.

$$\sum_0^n (k-nx)^2 \binom{n}{k} x^k(1-x)^{n-k} = nx(1-x) \leq \frac{n}{4}; \qquad n \geq 1.$$

If we write $(k-nx)^2$ as the sum of four terms $k(k-1)+k+n^2x^2-2nxk$ and simplify the four sums using the earlier equations. The last inequality is clear becaue for $x \in [0,1]$ we have $\sqrt{x(1-x)} \leq [x+(1-x)]/2$.

We now define the Bernstein polynomials associated with a continuous function f on the interval $[0,1]$.

$$P_n(x) = \sum_0^n f(k/n)\binom{n}{k} x^k(1-x)^{n-k}; \quad x \in [0,1]; \quad n \geq 1.$$

This is clearly a polynomial in x. The values of the function f at certain points appear in the definition of these polynomials.

The motivation for looking at such polynomials occurs from Probability theory. Roughly, the reason why this polynomial should be close to f is the following. The binomial coefficients $\binom{k}{n}x^k(1-x)^{n-k}$ increase as k increases from zero to near nx, reaching a maximum value near nx and then start decreasing. Also these coefficients are nearly zero at the tails. Thus in the above sum, $f(k/n)$ gets maximum weight whenever k is near nx. In other words values of f near x gets high weight in the above averaging and values away from x get weight very close to zero. Thus the average is close to $f(x)$.

To prove the theorem, fix $\epsilon > 0$. We show an N such that P_N satisfies the requirement. This is done as follows. First fix $\delta > 0$ so that $|f(x)-f(y)| < \epsilon/2$ whenever $|x-y| < \delta$, possible by uniform continuity of f. Fix a number C so that $|f(x)| < C$ for every $x \in [0,1]$, possible because continuous function on a closed bounded interval is bounded. Finally, fix integer $N > (C/\delta^2\epsilon)$. This will do. Actually, we show, $|f(x)-P_n(x)| < \epsilon$ for every $n \geq N$ and every $x \in [0,1]$.

From (♠) we see

$$f(x) = \sum_0^n \binom{n}{k} f(x)x^k(1-x)^{n-k},$$

so that

$$|f(x)-P_n(x)| = |\sum_k [f(x)-f(k/n)]\binom{n}{k} x^k(1-x)^{n-k}|$$

$$\leq \sum_k |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k}$$
$$= \sum{}' |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k}$$
$$+ \sum{}'' |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k}$$

Where $\sum'$ is sum over $\{k : |x - \frac{k}{n}| \leq \delta\}$. But by choice of δ this sum is at most $(\epsilon/2)$ times the sum of the binomial coefficients. Hence $\sum' \leq \epsilon/2$.

The sum $\sum''$ is over $\{k : |x - \frac{k}{n}| > \delta\}$. But by choice of C this sum is at most $2C$ times sum of the binomial coefficients. For every k in this sum, we have $|nx - k| > n\delta$ so that

$$\sum{}'' |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k}$$
$$\leq 2C \sum{}'' \binom{n}{k} x^k (1-x)^{n-k}$$
$$\leq 2C \sum{}'' \frac{(k-nx)^2}{n^2\delta^2} \binom{n}{k} x^k (1-x)^{n-k}$$
$$\leq \frac{2C}{n^2\delta^2} \frac{n}{4} = \frac{1}{2} \frac{C}{n\delta^2} = \frac{\epsilon}{2}.$$

This completes the proof of the theorem.

Why did we take the interval $[0, 1]$? Just to conveniently describe the polynomials. Any closed bounded interval is as good. More precisely, let f be a continuous function on a closed bounded interval $[a, b]$ and $\epsilon > 0$. Then there is a polynomial P such that

$$\sup_{x \in [a,b]} |f(x) - P(x)| < \epsilon.$$

This is observed by changing the action to the unit interval and coming back as follows. define $g(x) = f(a + [b-a]x)$ on $[0, 1]$. This is continuous, get a polynomial Q so that $|g(x) - Q(x)| < \epsilon$ for all $x \in [0, 1]$. Then $P(x) = Q([x-a]/[b-a])$ is a polynomial and serves our purpose. in fact $f(x) - P(x) = g([x-a]/[b-a]) - Q([x-a]/[b-a])$.

Can we do on the real line? That is, given a continuous function f on R and $\epsilon > 0$, can we find a polynomial P so that $|f(x) - P(x)| < \epsilon$ for all

$x \in R$. This is false in general. In fact take $f(x) = \sin x$ and $\epsilon = 1/4$. if you take a constant polynomial, it will not do because f takes values zero as well as one. If you take a non-constant polynomial, then it is not bounded.

L'Hopital revisited:

We shall discuss two issues connected with L'Hopital rule. Recall, it tells us that if $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$, and if $f'(x)/g'(x) \to \alpha$, then $f(x)/g(x) \to \alpha$. This is the $0/0$ form. We shall now show that the same result holds even for ∞/∞ form.

L'Hopital's rule: Let f and g be two differentiable functions on (a, b) such that $\lim f(x)$ as well as $\lim g(x)$ are ∞ as $x \to a$; and $g'(x) \neq 0$ on (a, b) and $\lim f'(x)/g'(x) \to \alpha$ as $x \to a$. Then $\lim f(x)/g(x) \to \alpha$ as $x \to a$.

The proof proceeds along the same lines as the $0/0$ case, but a little more involved.

We treat the case $\alpha \in R$. Other cases, namely, $\alpha = \infty$ and $\alpha = -\infty$ are similar. So let $\epsilon > 0$ be given. Need to show a number $\delta > 0$ so that

$$x \in (a, a + \delta) \Rightarrow |\frac{f(x)}{g(x)} - \alpha| < \epsilon.$$

As earlier, using the hypothesis, fix $\delta_1 > 0$

$$x \in (a, a + \delta_1) \Rightarrow \alpha - \frac{\epsilon}{4} < \frac{f'(x)}{g'(x)} < \alpha + \frac{\epsilon}{4}.$$

Let us fix a number y with $a < y < a + \delta_1$ (sort of a reference point and will not be changed in our calculations from now). The generalised MVT implies

$$a < x < y \Rightarrow \alpha - \frac{\epsilon}{4} < \frac{f(x) - f(y)}{g(x) - g(y)} < \alpha + \frac{\epsilon}{4}.$$

Take δ_2 so that $0 < \delta_2 < \delta_1$ and $g(x) > \max\{g(y), 0\}$ for $x \in (a, a + \delta_2)$. This is possible because $g(x) \to \infty$ as $x \to a$. Now let us take any $x \in (a, a + \delta_2)$ and multiply the above inequality by the positive number $[g(x) - g(y)]/g(x)$. Thus for $a < x < a + \delta_2$ we have

$$\left(\alpha - \frac{\epsilon}{4}\right) \frac{g(x) - g(y)}{g(x)} + \frac{f(y)}{g(x)} < \frac{f(x)}{g(x)} < \left(\alpha + \frac{\epsilon}{4}\right) \frac{g(x) - g(y)}{g(x)} + \frac{f(y)}{g(x)}.$$

Let us observe that $[g(x) - g(y)]/g(x) \to 1$ as $x \to a$ because y is fixed and $g(x) \to \infty$. Thus we can choose δ_3 so that $0 < \delta_3 < \delta_2$ and for $a < x < a + \delta_3$

$$\left(\alpha - \frac{\epsilon}{4}\right) \frac{g(x) - g(y)}{g(x)} > \alpha - \frac{\epsilon}{2}; \qquad \left(\alpha + \frac{\epsilon}{4}\right) \frac{g(x) - g(y)}{g(x)} < \alpha + \frac{\epsilon}{2}.$$

Thus for

$$a < x < a + \delta_3 \Rightarrow \left(\alpha - \frac{\epsilon}{2}\right) + \frac{f(y)}{g(x)} < \frac{f(x)}{g(x)} < \left(\alpha + \frac{\epsilon}{2}\right) + \frac{f(y)}{g(x)}.$$

Since $f(y)/g(x) \to 0$ as $x \to a$, choose δ_4 so that $0 < \delta_4 < \delta_3$ and for $x \in (a, a + \delta_4)$ this ratio is between $-\epsilon/2$ and $+\epsilon/2$. Thus we have

$$a < x < a + \delta_4 \Rightarrow \alpha - \epsilon < \frac{f(x)}{g(x)} < \alpha + \epsilon.$$

This completes proof of the rule.

The second issue related to L'Hopital's rule is the following. Let us again consider the 0/0 case. What if f' and g' also converge to zero at a? The answer is that we can try second derivatives and so on.

Fact: Suppose f and g are $(n-1)$-times continuously differentiable on an interval $[a, b]$ and f as well as g and all their first $(n-1)$ derivatives are zero at a. Suppose $f^{(n)}$ and $g^{(n)}$ exist in the open interval; $g^{(n)}(x) \neq 0$ for all $x \in (a, b)$ and $f^{(n)}(x)/g^{(n)}(x) \to \alpha$ as $x \downarrow a$. then $f(x)/g(x) \to \alpha$ as $x \downarrow a$.

As suggested by Pranav, you can use the earlier case repeatedly as follows. First observe that $g^{(n)}$ never takes the value zero, so that $g^{(n-1)}(x)$ can be zero for at most one value of x in (a, b) — use mean value theorem. Say it is nonzero in (a, b_1). Repeat this argument to see $g^{(n-2)}$ is non-zero in (a, b_2) and so on; finally getting an interval (a, β) where all these derivatives of g are non-zero and consider only this interval in what follows. This is alright because we are interested as $x \to a$.

Applying earlier version of L'Hopital successively to $f^{(k)}; g^{(k)}$ with $k = n-1, n-2, \cdots 1$ deduce

$$\lim_{x \downarrow a} \frac{f^{(n)}(x)}{g^{(n)}(x)} = \lim_{x \downarrow a} \frac{f^{(n-1)}(x)}{g^{(n-1)}(x)} = \lim_{x \downarrow a} \frac{f^{(n-2)}(x)}{g^{(n-2)}(x)} = \cdots = \lim_{x \downarrow a} \frac{f(x)}{g(x)}.$$

Here is another method when α is finite. This depends on a generalization of MVT with higher order derivatives or Taylor for two functions.

Fact: f, g are $n-1$ times continuously differentiable in the interval $[a,b]$ and $f^{(n)}$, $g^{(n)}$ exist in the interval (a,b). There is a number $\theta \in (a,b)$ such that

$$\left[f(b) - \sum_0^{n-1} f^{(k)}(a)\frac{(b-a)^k}{k!}\right] g^{(n)}(\theta) = \left[g(b) - \sum_0^{n-1} g^{(k)}(a)\frac{(b-a)^k}{k!}\right] f^{(n)}(\theta).$$

Proof consists of applying the earlier version to the following functions.

$$F(x) = f(x) + \sum_1^{n-1} \frac{f^{(k)}(x)}{k!}(b-x)^k.$$

$$G(x) = g(x) + \sum_1^{n-1} \frac{g^{(k)}(x)}{k!}(b-x)^k.$$

These are continuous on $[a,b]$ and differentiable in (a,b) so there is $\theta \in (a,b)$ such that

$$[F(b) - F(a)]G'(\theta) = [G(b) - G(a)F'(\theta). \qquad (\clubsuit)$$

$$F(b) - F(a) = f(b) + 0 - f(a) - \sum_1^{n-1} \frac{f^{(k)}(a)}{k!}(b-a)^k$$

$$= f(b) - \sum_0^{n-1} f^{(k)}(a)\frac{(b-a)^k}{k!} \qquad (*)$$

For $x \in (a,b)$

$$G'(x) = g'(x) + \sum_1^{n-1} \frac{g^{(k+1)}(x)}{k!}(b-x)^k - \sum_1^{n-1} \frac{g^{(k)}(x)}{(k-1)!}(b-x)^{k-1}$$

$$= g^{(n)}(x)\frac{(b-a)^{n-1}}{(n-1)!}. \qquad (**)$$

Similarly,

$$G(b) - G(a) = g(b) - \sum_0^{n-1} g^{(k)}(a)\frac{(b-a)^k}{k!} \qquad (\dagger)$$

and

$$F'(x) = f^{(n)}(x)\frac{(b-a)^{n-1}}{(n-1)!}. \qquad (\dagger\dagger)$$

Substituting $(*), (**), (\dagger), (\dagger\dagger)$ in $(\clubsuit)$ we get the result.

Now the proof of L'Hopital goes exactly as in the earlier case. Fix $\epsilon > 0$. Need to show $\delta > 0$ so that

$$a < x < a + \delta \Rightarrow \alpha - \epsilon < \frac{f(x)}{g(x)} < \alpha + \epsilon.$$

Fix, using hypothesis, $\delta > 0$ so that

$$a < x < a + \delta \Rightarrow \alpha - \epsilon < \frac{f^{(n)}(x)}{g^{(n)}(x)} < \alpha + \epsilon.$$

The same δ would do. take any $a < y < x < a + \delta$ apply the above MVT on the interval $[y, x]$ and let $y \downarrow a$. This completes the alternative proof.

Newton's algorithm for zero:

We shall discuss just one inteersting computational application of the Taylor formula. It is to find a zero of a function. Shall present the most primitive version of an algorithm Newton discovered.

Suppose we have a function $f : R \to R$. Let z be a zero of f, that is, $f(z) = 0$. Find an algorithm to calculate z with required degree of accuracy, if we know roughly where it is located.

Start with a point z_0. This is your initial or starting approximation to z (and hence, in general, you need to start close to z). Draw tangent to the curve, graph of f, at $(z_0, f(z_0))$. Suppose it cuts the x-axis at z_1 (so you need to assume that tangent is not parallel to the x-axis). This is your first approximation to z. Then draw tangent to the curve at $(z_1, f(z_1))$ Suppose that it cuts the x-axis at z_2. This is your second approximation. Continue the process. The hope is you are heading towards to z, the actual zero.

What makes us hope so? Well, as is always the case, look at some examples. For instance consider the curve $f(x) = x^2 - 2$ and take $z_0 = 1$ and try. Take the curve $f(x) = x^3 - 2$ and try again. of course, the fact that ti works in the examples we have seen is not good enough to believe that this is always true. In fact it is not always true that these numbers z_n so obtained converge to the actual zero. But, under fairly general conditions they do converge.

Let us see what the algiorithm says. The tangent at $(z_0, f(z_0))$ to the curve is given by

$$y - f(z_0) = f'(z_0)(x - z_0).$$

To find the point of intersection with the x-axis, set $y = 0$ to see

$$z_1 = z_0 - \frac{f(z_0)}{f'(z_0)}.$$

To get z_2 you repeat the same formula above. Here is the precise fact.

Suppose that f is a twice differentiable function in an open interval I. Suppose that $f'(x)$ is never zero in this interval. Assume that

$$\frac{\sup |f''(x)|}{\inf |f'(x)|} = \alpha < \infty.$$

Let z be a point in I with $f(z) = 0$.
Start with a point $z_0 \in I$ with $\alpha|z - z_0| < 1$. Define recursively,

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}.$$

Then each of the points z_n is in the interval I and $z_n \to z$.

Here is the proof. Suppose $z_n \in I$. By Taylor expansion around z_n, there is a point θ between z_n and z such that

$$f(z) = f(z_n) + f'(z_n)(z - z_n) + \frac{1}{2} f''(\theta)(z - z_n)^2.$$

Since $f(z) = 0$ we gets

$$f(z_n) = -f'(z_n)(z - z_n) - \frac{1}{2} f''(\theta)(z - z_n)^2.$$

Hence

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} = z_n + (z - z_n) + \frac{1}{2} \frac{f''(\theta)}{f'(z_n)}(z - z_n)^2.$$

$$|z_{n+1} - z| = |\frac{1}{2} \frac{f''(\theta)}{f'(z_n)}(z - z_n)^2.| \leq \frac{1}{2}\alpha|z_n - z|^2$$

In particular

$$|z_1 - z| \leq \frac{1}{2}\{\alpha|z_0 - z|\}|z_0 - z| < \frac{1}{2}|z_0 - z|.$$

showing that $z_1 \in I$ and in fact closer to z than z_0. In particular $\alpha|z_1 - z| < 1$. Thus

$$|z_2 - z| \leq \frac{1}{2}\{\alpha|z_1 - z|\}|z_1 - z| < \frac{1}{2}|z_1 - z|. \leq \frac{1}{2^2}|z_0 - z|.$$

By induction, you can conclude the following for each n:

$$z_n \in I; \qquad \alpha |z_{n-1} - z| < 1; \qquad |z_n - z| \le \left(\frac{1}{2}\right)^n |z_0 - z|.$$

This completes proof of the assertion made. In fact the convergence of the approximations is 'geometric'.

Integration:

Here is the problem, familiar from high school. How do you calculate areas. of course, if we have a rectangular region, we some how seem sure and agreed upon that its area is product of lengths of sides. Thus areas of rectangles are taken as known. Of course, from this we some how built up several other areas, for example area of a circle of radius r equals πr^2. We also know how to calculate areas of triangles and some other figures like parallelogram, trapezium etc.

Let us consider a function f defined on the interval $[0,1]$. As a first step, let us assume that the function takes only non-negative values, so that the graph of the function is above the x-axis. We also assume that the function is bounded. Consider the region under the curve, bounded below by the x-axis. On the sides it is bounded by the vertical lines at $x = 0$ and $x = 1$. Thus to the left, the region is bounded by the y-axis and to the right, it is bounded by the vertical line at $x = 1$.

Analytically, it is the region $\{(x, y) : 0 \le x \le 1, 0 \le y \le f(x)\}$.

How nice if the function is a constant $f \equiv c > 0$. Then the graph of f is just a flat horizontal line at height c and thus the region is just a rectangle with sides of lengths one and c. Its area is c. Suppose the function is piece-wise constant, say, on $[0, 1/4]$ value of f is α; on $(1/4, 1/3)$ value of f is β and finally on $[1/3, 1]$ value of f is γ. Then the region under consideration consists of three rectangles and so the 'total' area of the region equals $\alpha(1/4) + \beta(1/12) + \gamma(2/3)$.

If the function is not piece-wise constant, then the region is, of course, not made up of finitely many rectangles. Let us denote the area by A. The plan is to find a number which bounds A from below and to find a number which bounds A from above. For instance, suppose that there are reasons to believe that the unknown area A must at least be 3 and also can be at most 3. Then you must agree that the area must equal 3, neither more nor less.

Of course, if you can only conclude that the area must at least be 3 and can not exceed 4, then we are sure that it must be a number between 3 and 4; but not yet sure what exactly it should equal.

If you are trying to see a lower bound for the area, the best way is to fit in non-overlaping rectangles in our region and take their total area. this can be done in several ways. what is the most efficient way of doing it? here is what Riemann's idea is. Break up the interval $[0, 1]$ into finitely many intervals, say $[0, a_1]$, $[a_1, a_2]$ $\cdots$ $[a_{55}, 1]$ — 56 pieces. On each piece, let us calculate the infimum of the function. Let them be $m_0, m_1, \cdots, m_{55}$. Consider the rectangles with base $[a_i, a_{i+1}]$ and height m_i for $i = 0, 1, 2, \cdots, 55$. Clearly these rectangles are non-overlapping (have common sides) and are all within the region of our interest. Thus the area A must be at least sum of the areas of theses rectangles.

For every break up of the interval $[0, 1]$ we calculate a number, namely, sum of the areas of the rectangles constructed as above, by taking infimum of the function in each of the intervals. Let L be the set of numbers so obtained. Thus, we must agree to the following. Whatever be the area A, we must have $\alpha \leq A$ for each $\alpha \in L$. ln other words A is an upper bound for the set L. Let l be the least upper bound of the set L, that is, supremum of L. Thus we have $l \leq A$.

Let us use the same notation as in the earlier para, and now let us consider the supremum of the function in each of the intervals, say, $M_0, M_1, \cdots, M_{55}$. Consider the rectangles with base $[a_i, a_{i+1}]$ and height M_i for $i = 0, 1, 2, \cdots, 55$. Clearly these rectangles are non-overlapping (have common sides) and all the rectangles put together include the region of our interest. Thus the area A can at most be sum of the areas of these rectangles.

For every break up of the interval $[0, 1]$ we calculate a number, namely, sum of the areas of the rectangles constructed as above, by taking supremum of the function in each of the intervals. Let U be the set of numbers so obtained. Thus, we must agree to the following. Whatever be the area A, we must have $A \leq \alpha$ for each $\alpha \in U$. ln other words A is a lower bound for the set U. Let u be the infimum or the glb of the set U. Thus we have $A \leq u$.

Thus we have calculated two numbers l and u and we are sure that $l \leq A \leq u$. In particular, if we are lucky and it so turns out that $l = u$ then this *must* be the area of the region and there is nothing for us to decide. What is amazing is that this equality holds in many cases. Thus this intuitive algorithm leads to an answer for the concept of area in many situations.

Of course, it does not lead to an answer in many other cases. For example, if our function f is given by: $f(x)$ equals one or zero according as x is rational or not. Then L consists of one number, namely 0; whereas U consists of one number 1. thus $l = 0$ and $u = 1$, unequal.

We shall now start with definitions execute the above idea. In what follows $[a, b]$ is a closed bounded interval. By a partition of this interval we mean a finite set of points $a = a_0 < a_1 < a_2 < \cdots, a_k = b$. We denote partition by P. Thus partition is just a finite subset of $[a, b]$ and $a \in P$ and $b \in P$. Of course a set consists of only points and there is nothing like first or second element in the set. When we think of a partition, we keep the order also in mind when we picturize it. A partition as above breaks the interval into finitely many intervals, namely, $[a, a_1]$, $[a_1, a_2]$, $\cdots$, $[a_{k-1}, b]$. Sometimes we refer to these intervals as intervals of the partition.

If P_1 and P_2 are two partitions, we say P_2 is finer than P_1 if every point of P_1 is also in P_2, that is, $P_1 \subset P_2$. We write $P_1 \leq P_2$. Thus, for example if the interval is $[0, 1]$ then

$$\{0, 1/2, 1\}; \quad \{0, 1/4, 1/3, 1/2, 1\}; \quad \{0, 1/8, 3/8, 1\}$$

are all partitions. The second one is a refinement of the first one. The third one is not comparable to the other two. However we can think of a partition that refines both the second and third, namely

$$\{0, 1/8, 1/4, 1/3, 3/8, 1/2, 1\}.$$

In general refinement of two partitions P_1 and P_2 is just $P_1 \cup P_2$, of course, you need to arrange the points in increasing order when you picturize. This is denoted by $P_1 \vee P_2$

Alternatively, you can define a partition as a finite increasing subset of $[a, b]$ by bringing in the order also as part of the definition. Then of course common refinement would consist of increasing arrangement of the points of both taken together.

Now let f be a bounded real valued function on $[a, b]$. For a partition $P = \{a = a_0 < a_1 < a_2 < \cdots < a_k = b\}$, we define

$$U(P, f) = \sum_0^{k-1} M_i(a_{i+1} - a_i); \qquad M_i = \sup\{f(x) : a_i \leq x \leq a_{i+1}\}.$$

$$L(P,f) = \sum_{0}^{k-1} m_i(a_{i+1} - a_i); \qquad m_i = \inf\{f(x) : a_i \le x \le a_{i+1}\}.$$

$U(P,f)$ is the upper Riemann sum for the partition P and $L(P,f)$ is the lower Riemann sum for the partition P. We put

$$U(f) = \inf_P U(P,f); \qquad L(f) = \sup_P L(P,f).$$

In the above, the inf and sup is over all the partitions of the interval $[a,b]$.

We say that f is *Riemann integrable* if $U(f) = L(f)$ and then define integral of f as this common value $U(f) = L(f)$. The integral is denoted $\int_a^b f(x)dx$. Convince yourself that this is precisely what we thought at the beginning.

Here are some facts regarding the Riemann sums and integrals.

1: For any partition P, $L(P,f) \le U(P,f)$
This follows from the fact $m_i \le M_i$ for each i.

2. If $P_1 \le P_2$, then $L(P_1,f) \le L(P_2,f)$ and $U(P_1,f) \ge U(P_2,f)$.

First suppose that P_2 has only one extra point than P_1; say

$$P_1 = \{a = a_0 < a_1 < \cdots < a_k = b\}.$$

Suppose P_2 has one extra point α; $a_j < \alpha < a_{j+1}$. Let

$$m' = \inf\{f(x) : a_j \le x \le \alpha\}; \quad m'' = \inf\{f(x) : \alpha \le x \le a_{j+1}\}.$$

Then m_j being $\inf\{f(x) : a_j \le x \le a_{j+1}\}$ we see $m_j \le m'$ and $m_j \le m''$; as the set gets larger, inf gets smaller. Thus

$$\begin{aligned} m_j(a_{j+1} - a_j) &= m_j(\alpha - a_j) + m_j(a_{j+1} - \alpha) \\ &\le m'(\alpha - a_j) + m''(a_{j+1} - \alpha). \end{aligned}$$

Observe that the only difference between $L(P_1,f)$ and $L(P_2,f)$ is the following. The term $m_j(a_{j+1} - a_j)$ appearing in $L(P_1,f)$ is replaced by the right side above in $L(P_2,f)$. Thus the above inequality shows that $L(P_1,f) \le L(P_2,f)$. Similarly if

$$M' = \sup\{f(x) : a_j \le x \le \alpha\}; \quad M'' = \sup\{f(x) : \alpha \le x \le a_{j+1}\}.$$

then we see $M' \leq M_j$ and $M'' \leq M_j$ leading to the inequality $U(P_2, f) \leq U(P_1, f)$.

Clearly, by induction, the inequalities follow if P_2 has k extra points than P_1; $k = 1, 2, \cdots$ This completes the proof.

3. For any two partitions P_1 and P_2, we have $L(P_1, f) \leq U(P_2, f)$.
In fact if $P = P_1 \vee P_2$, their refinement, we see using the two facts above

$$L(P_1, f) \leq L(P, f) \leq U(P, f) \leq U(P_2, f).$$

4. $L(f) \leq U(f)$.
The earlier fact says that every number $L(P, f)$ is a lower bound for the set $S = \{U(Q, f) : Q \text{ partition of } [a, b]\}$. Thus $L(f)$ is a lower bound for S. But $U(f)$ is the glb of S. Hence $L(f) \leq U(f)$.

5. If $\alpha \leq f(x) \leq \beta$ for all $x \in [a, b]$, then for each partition P,
$\alpha(b - a) \leq L(P, f)$ and $U(P, f) \leq \beta(b - a)$.

Just note that $\alpha \leq m_j \leq M_j \leq \beta$ for each j.

6. If $\alpha \leq f(x) \leq \beta$ for all $x \in [a, b]$, then

$$\alpha(b - a) \leq L(f) \leq U(f) \leq \beta(b - a).$$

This follows from the above.

7. Every continuous function on $[a, b]$ is integrable.
We need to show that $L(f) = U(f)$. Since $L(f) \leq U(f)$ always, we only need to show $U(f) \leq L(f)$. Fix $\epsilon > 0$. We show a partition P so that $U(P, f) - L(P, f) < \epsilon$ Then it follows that

$$U(f) - L(f) \leq U(P, f) - L(P, f) \leq \epsilon$$

Since $\epsilon > 0$ is arbitrary, it follows that $U(f) \leq L(f)$ as required.

Use uniform continuity of f on $[a, b]$ to get $\delta > 0$ so that

$$x, y \in [a, b]; |x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{b - a}.$$

Now take any partition

$$P = \{a = a_0 < a_1 < \cdots < a_k = b\}$$

such that each interval of the partition has length smaller than δ. There are such partitions. For example, $a, a + (\delta/2), a + 2(\delta/2), \cdots$. Of course, when $k(\delta/2)$ exceeds b, you stop and take the last point as b. If x and y are in $[a_j, a_{j+1}]$ we see $|f(x) - f(y)| < \epsilon/(b-a)$ and hence $M_j - m_j < \epsilon/(b-a)$. In fact, f being continuous, you can take x and y to be the points in the interval $[a_j, a_{j+1}]$ where the max and min are attained. Thus

$$U(P, f) - L(P, f) = \sum(M_j - m_j)(a_{j+1} - a_j)$$

$$\leq \frac{\epsilon}{b-a} \sum(a_{j+1} - a_j) = \epsilon.$$

Thus we see that a large class of functions, namely continuous functions, are integrable.