CHAIN RULE, u-SUBSTITUTION, SYMMETRY, MEAN VALUE THEOREM

MATH 152, SECTION 55 (VIPUL NAIK)

Corresponding material in the book: Section 5.6, 5.7, 5.8, 5.9.
Difficulty level: Hard.
What students should definitely get: The idea of using differentiation rules to determine antiderivative, the application of the chain rule to indefinite integration, and the idea of the u-substitution. The application of the u-substitution to definite integrals, the idea that definite integrals can be computed in light of certain kinds of symmetry even without computing an antiderivative. Bounding definite integrals via other definite integrals. The mean value theorem for integrals.
What students should eventually get: A grasp and clear memory of all the rules for computing definite integrals for functions with a certain kind of symmetry. The subtleties of u-substitutions.

Executive summary

0.1. Reversing the chain rule. Actions ...

(1) The chain rule states that $(f \circ g)' = (f' \circ g) \cdot g'$.
(2) Some integrations require us to reverse the chain rule. For this, we need to realize the integrand that we have in the form of the right-hand side of the chain rule.
(3) The first step usually is to find the correct function g, which is the *inner function* of the composition, then to adjust constants suitably so that the remaining term is g', and then figure out what f' is. Finally, we find an antiderivative for f', which we can call f, and then compute $f \circ g$.
(4) A slight variant of this method (which is essentially the same) is the substitution method, where we identify g just as before, try to spot g' in the integrand as before, and then put $u = g(x)$ and rewrite the integral in terms of u.

0.2. u-substitutions and transformations. Words ... (try to recall the numerical formulations)

(1) When doing the u-substitution for definite integrals, we transform the upper and lower limits of integration by the u-function.
(2) Note that the u-substitution is valid only when the u-function is well-defined on the entire interval of integration.
(3) The integral of a translate of a function is the integral of a function with the interval of integration suitably translated.
(4) The integral of a multiplicative transform of a function is the integral of the function with the interval of integration transformed by the same multiplicative factor, scaled by that multiplicative factor.

0.3. Symmetry and integration. Words ...

(1) If a function is continuous and even, its integral on $[-a, 0]$ equals its integral on $[0, a]$. More generally, its integrals on any two segments that are reflections of each other about the origin are equal. As a corollary, the integral on $[-a, a]$ is twice the integral on $[0, a]$.
(2) If a function is continuous and odd, its integral on $[-a, 0]$ is the negative of its integral on $[0, a]$. More generally, its integrals on any two segments that are reflections of each other about the origin are negatives of each other. As a corollary, the integral on $[-a, a]$ is zero.
(3) If a function is continuous and has mirror symmetry about the line $x = c$, its integral on $[c - h, c]$ equals its integral on $[c, c + h]$.
(4) If a function is continuous and has half-turn symmetry about $(c, f(c))$, its integral on any interval of the form $[c - h, c + h]$ is $2hf(c)$. Basically, all the variation about $f(c)$ *cancels out* and the *average value* is $f(c)$.

(5) Suppose f is continuous and periodic with period h and F is an antiderivative of f. The integral of f over any interval of length h is constant. Thus, $F(x+h) - F(x)$ is the same constant for all x. (We saw this fact long ago, without proof).
(6) The constant mentioned above is zero iff F is periodic, i.e., f has a periodic antiderivative.
(7) There is thus a well-defined *average value* of a continuous periodic function on a period. This is also the average value of the same periodic function on any interval whose length is a nonzero integer multiple of the period. This is also the limit of the average value over very large intervals.

Actions...

(1) All this even-odd-periodic stuff is useful for trivializing some integral calculations without computing antiderivatives. This is more than an idle observation, since in a lot of real-world situations, we get functions that have some obvious symmetry, even though we know very little about the concrete form of the functions. We use this obvious symmetry to compute the integral.
(2) Even if the whole integrand does not succumb to the lure of symmetry, it may be that the integrand can be written as (something nice and symmetric) + (something computable). The (nice and symmetric) part is then tackled using the ideas of symmetry, and the computable part is computed.

0.4. **Mean-value theorem.** Words ...

(1) The *average value*, or *mean value*, of a continuous function on an interval is the quotient of the integral of the function on the interval by the length of the interval.
(2) The mean value theorem for integrals says that a continuous function must attain its mean value somewhere on the interior of the interval.
(3) For periodic functions, the mean value over any interval whose length is a multiple of the period is the same. Also, the mean value over a very large interval approaches this value.

1. Philosophical remarks on hardness

1.1. **A fundamental asymmetry between differentiation and integration.** A little while back in the course, we saw how to differentiate functions. In order to carry out differentation, we learned how to differentiate all the basic building block functions (the polynomial functions and the sine and cosine functions) and then we learned a bunch of rules that allowed us to differentiate any function built from these elementary functions using either function composition or pointwise combination.

This means that for any function built from the elementary functions, if we know how to write it, we know how to compute its derivative. The strategy is to keep breaking down the task using the rules for combination and composition until we get to differentiating the elementary functions, fo which we have formulas.

The analogue is *not* true for finding antiderivatives. In other words, there is no foolproof procedure to break down operations such as combination and composition and ultimately reduce the problem to computing antiderivatives of the basic building blocks. Thus, *even though there are formulas* for the antiderivatives of all the basic building blocks, there exist functions constructed from these that do not have antiderivatives that can be written as elementary functions.

One important point should be made here. Just because the antiderivative of f cannot be expressed as an elementary function (or, it can but we're not able to determine that elementary function) does *not* mean that the antiderivative does not exist. Rather, it means that the existing pool of functions that we have is not large enough to contain that function, and we may need to introduce new classes of functions.

1.2. **Dealing with failure, getting used to it.** Examples of functions that do not have antiderivatives in the classes of functions we have seen so far include $1/(x^2+1)$, $1/\sqrt{x^2+1}$, $1/x$, and $1/\sqrt{x^2-1}$. Later in the course, we shall introduce new classes of functions, and it turns out that these functions are integrable within those new classes of functions. (The new classes include logarithmic functions and inverse trigonometric functions). Functions such as $1/\sqrt{x^3+2x+7}$ cannot be integrated even in this larger collection of functions – to integrate these functions, we would need to introduce *elliptic functions* and *inverse elliptic functions* which are an analogue of trigonometric and inverse trigonometric functions. We won't formally introduce those functions in the 150s, and you probably will not see them ever in a formal way.

Similarly, the functions $\sin(x^2)$ and $(\sin x)/x$ do not have indefinite integrals expressible in our current vocabulary (the integral of the latter is particularly important and is called the *sine integral*, even though it has no easy expression). When we later introduce the logarithmic function, we will see that $1/\log x$ has no antiderivative in the classes of functions we are dealing with, though one of its antiderivatives, called the *logarithmic integral*, is extremely important in number theory and in the distribution of prime numbers.

When we later introduce the exponential function, we shall see that e^{-x^2} has no antiderivative expressible in terms of elementary functions. However, the antiderivative of e^{-x^2} is *extremely important* in statistics, since e^{-x^2} corresponds to the shape of the Gaussian or normal distribution (a shape often called a *Bell curve*) and its integral measures the area under the curve for such a distribution. The integral is so important in statistics that there are tables of the values of the definite integral from 0 to a for different numerical values of a. These tables can be used to calculate the definite integral between any two points. As an interesting aside, it is true that the integral of e^{-x^2} over the entire real line (a concept we will see later) is $\sqrt{\pi}$.

In general, the use of antiderivatives and indefinite integration is a powerful tool in performing definite integration. Recall that if F is an antiderivative for f, then $\int_a^b f(x)\,dx = F(b) - F(a)$. So, to integrate f between limits, all we need to do is find an antiderivative, evaluate it at the limits, and subtract. However, there are three problems that we encounter as soon as we start trying this approach for nontrivial functions:

(1) An antiderivative may not even exist within the class of functions that we are familiar with. In other words, we may need to define and introduce new classes of functions to fit in the antiderivative. This is not very helpful for computational purposes.
(2) Even if the antiderivative exists, it may require considerable ingenuity to find it. This is because there is no clear and short step-by-step reductive algorithm to find an antiderivative. This is in sharp contrast with the situation for derivatives, where we can reduce step by step.
(3) Even if we successfully calculate the antiderivative, it may not be much use to us computationally if we cannot evaluate the antiderivative at the two endpoints. This is more of a problem when dealing with functions that involve trigonometric functions (and inverse trigonometric, exponential, and logarithmic functions – new classes of functions you have been sheltered from so far).

In all these cases, one tool still remains at our disposal – the *back-to-basics* definition of the definite integral using upper sums and lower sums. This definition can usually allow us to quickly obtain crude upper bounds and crude lower bounds. Such bounds are not as good as an exact answer but they may be good enough.

2. Breaking the differentiation code: reverse engineering

2.1. **Recalling the rules for differentiation.** Our next stop is the rules for differentiation. Broadly, our strategy for computing antiderivatives is *working backward*: starting from rules that we know for differentiation and trying to guess what the antiderivative must have been so that differentiating it gives the function we have at hand.

We have already seen the rules for sums and scalar multiples for differentiation. In technical terminology, we say that the antiderivative is *linear* – the antiderivative of the sum is the sum of the antiderivatives, and the antiderivative of a scalar multiple is the same scalar multiple of the antiderivative (I'm being imprecise by using *the*, but you should get the idea).

There are two other rules for differentiation that are somewhat more complicated: the *product rule* and the *chain rule*. In this lecture, we concentrate on the chain rule. The product rule manifests itself in a technique, called *integration by parts*, that we will see next quarter.

2.2. **The chain rule.** Let's look at the chain rule.

$$(f \circ g)' = (f' \circ g) \cdot g'$$

Equivalently:

$$\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$$

In order to use the chain rule to integrate a function p, we need to do what's called *pattern matching* – we need to find functions f and g such that we can write $p = (f' \circ g) \cdot g'$. In some cases, the way the function is written is reasonably suggestive of what f and g are. In other cases, we need to do a little work. We look at some examples.

Consider:

$$\int \sin(\cos x) \sin x \, dx$$

Comparing this with the general expression, we see that we should have $g(x) = \cos x$, since cos is the *inner function* of the composition in this expression. If $g(x) = \cos x$, then we obtain $g'(x) = -\sin x$. We notice that the expression we have is $\sin x$, not $-\sin x$, so we put a negative sign on the outside, and obtain:

$$-\int \sin(\cos x)(-\sin x) \, dx$$

If $g(x) = \cos x$, then $g'(x) = -\sin x$. Looking again at the pattern we are trying to match, we see that we must have $f'(t) = \sin t$. We thus see that a possible candidate for $f(t)$ is $-\cos t$, since that is an antiderivative for sin. Using the chain rule in reverse, we thus obtain that $-(-\cos(\cos x)) = \cos(\cos x)$ is an antiderivative. The indefinite integral is thus:

$$\cos(\cos x) + C$$

where C is an arbitrary real constant.

It might be worthwhile to differentiate this and check that the derivative we get is the original integrand.

Here is another example. Consider:

$$\int (x^2 + 1)^{45} x \, dx$$

We can perform this integration by first computing $(x^2 + 1)^{45}$ as a polynomial, then multiplying each term by x, then integrating termwise. However, this is impractical. Instead, we try to use the chain rule.

The composite function of interest is $(x^2 + 1)^{45}$. This is a composite of the function $g(x) = x^2 + 1$ and the function $h(t) = t^{45}$. The derivative of g is $g'(x) = 2x$, which is twice of the expression we have (simply x). Thus, we need to multiply and divide by 2:

$$\frac{1}{2} \int (x^2 + 1)^{45} 2x \, dx$$

Now, we see that $f'(t) = h(t) = t^{45}$, so f is an antiderivative for that. We could take $f(t) = t^{46}/46$. The overall antiderivative then simplifies to:

$$\frac{1}{2} \frac{(x^2 + 1)^{46}}{46} + C = \frac{(x^2 + 1)^{46}}{92} + C$$

Let's look at another example:

$$\int x^3 \, dx$$

We already know that an antiderivative for this is $x^4/4$ and the general expression for the indefinite integral is $(x^4/4) + C$. We now see how this result can be obtained using the chain rule. We write $x^3 = x^2 \cdot x$. We then set $g(x) = x^2$, and $f'(t) = t$ (so $f(t) = t^2/2$), so that $x^2 = f'(g(x))$. We also have $g'(x) = 2x$, which is twice of x, so we get:

$$\frac{1}{2} \int (x^2) \cdot (2x) \, dx$$

The integral is thus:

$$\frac{1}{2} \frac{(x^2)^2}{2} + C = \frac{x^4}{4} + C$$

Let us look at one more example:

$$\int \frac{2x}{(x^2 + 1)^2} \, dx$$

Here, we notice that the derivative of $x^2 + 1$ is $2x$. Thus, we set $g(x) = x^2 + 1$. We obtain $g'(x) = 2x$. Also, we have $f'(g(x)) = 1/(x^2 + 1)^2$, so $f'(t) = 1/t^2$. Thus, $f(t)$ is an antiderivative for $1/t^2$, so we can set $f(t) = -1/t$. Plugging these in, we obtain that $f(g(x)) = -1/(x^2 + 1)$, so we obtain that the indefinite integral is:

$$\frac{-1}{x^2 + 1} + C$$

A glimpse into the u-substitution. One drawback of the approach outlined above for reverse-engineering the chain rule is that we have to do a lot of rough work and this becomes tedious for harder problems. An alternative way of presenting this, that makes things easier to handle in harder situations, is by using a substitution. Here, we identify $g(x)$ in the same way as we did earlier, and we then try to write our integral as:

$$\int h(g(x))g'(x)\,dx$$

The main difference is that instead of trying to find *a priori* a function f such that $f' = h$, we instead postpone that for later. We perform a substitution $u = g(x)$, whereby we replace $g'(x)dx$ with du, and obtain:

$$\int h(u)\,du$$

which we now proceed to integrate (which is essentially the same as finding an antiderivative for h, which is the function we called f). Finally, we substitute in $g(x)$ for u in the expression we obtain.

This seems like more steps. However, the main advantage is that one of the steps that we had to do as scratch work, namely finding f using the expression we have for f', is now done in the open. This is particularly useful if the function $h = f'$ is complicated and integrating it requires many steps.

There is also a slight variant of substitution for definite integrals. We now turn to that.

3. Magic of definite integrals with chain rule

3.1. **The u-substitution revisited.** Recall the u-substitution, which is a variant of the procedure to reverse the chain rule, but has the advantage that it breaks our work more clearly into two steps: first find g, then reduce the problem to a new problem that involves finding the antiderivative for h.

How do we use this to compute a definite integral? We can use the procedure outlined above to compute an antiderivative in terms of x, and then evaluate it between limits. For instance, consider:

$$\int_0^\pi \cos(\sin x)\cos x\,dx$$

We first try to compute the antiderivative:

$$\int \cos(\sin x)\cos x\,dx$$

Set $u = \sin x$. Then, the above integral becomes:

$$\int \cos u\,du$$

which is $\sin u$. Since $u = \sin x$, we obtain that $\sin(\sin x)$ is an antiderivative. The definite integral is thus $\sin(\sin \pi) - \sin(\sin 0) = 0$.

There is an alternative way of doing things, which involves *changing the limits of integration with each u-substitution.* The idea here is that every time we make a substitution of the form $u = g(x)$, we replace the lower and upper limits by their images under g. In other words, if the function is being integrated from a to b, the new function is being integrated from $g(a)$ to $g(b)$. In symbols:

$$\int_a^b h(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} h(u)\,du$$

The advantage of this is that after we find an antiderivative for h, say f, we do not need to compute the function $f \circ g$, i.e., we do not need to find an antiderivative for the original integrand. We simply evaluate the new antiderivative between the new limits $g(a)$ and $g(b)$.

The approach has other advantages, namely, in situations where it is difficult or impossible to get explicit expressions for antiderivatives, but a definite integral can be computed due to symmetry considerations or for other degenerate reasons. For instance, consider:

$$\int_0^\pi \cos(\sin x) \cos x \, dx$$

Set $u = \sin x$. The limits now become $\sin 0$ and $\sin \pi$, so the integral becomes:

$$\int_0^0 \cos u \, du$$

Note that with this u-substitution method, we do not even need to find an antiderivative for the integrand: we can straightaway compute that the integral is zero, because the upper and lower limits for integration coincide.

3.2. **Inequalities involving the definite integral.** We'll now review some of the properties of the definite integral that are discussed in Section 5.8 in the book. We begin with properties 5.8.1 – 5.8.4. These are fairly straightforward, and are expected from the notion of integral as a total value, or from the formal definition involving lower and upper sums. Note that by default, all integrals are over intervals of positive length, taken from left to right, i.e., the lower limit of the integral is strictly smaller than the upper limit of the integral.

1. The integral of a nonnegative continuous function is nonnegative. (5.8.1)
2. The integral of an everywhere positive function is positive. (5.8.2)
3. If $f(x) \le g(x)$ for all $x \in [a, b]$, then $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$. (5.8.3)
4. If $f(x) < g(x)$ for all $x \in [a, b]$, then $\int_a^b f(x)\,dx < \int_a^b g(x)\,dx$. (5.8.4)

These inequalities give us a new tool for bounding an integral from above and below. We now turn to that tool.

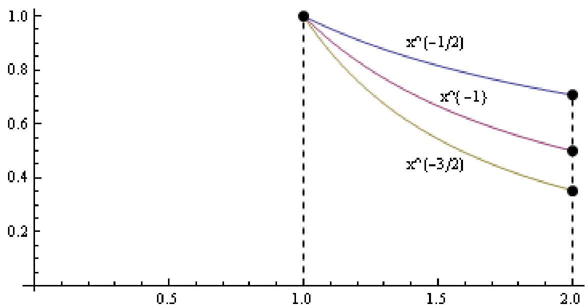
3.3. **Bounding an integral.** It is not always possible to find an explicit expression for an antiderivative. Hence, we cannot always compute the definite integral of a function via the antiderivative route. One strategy we had for overcoming this was the use of *upper and lower sums*. These sums, however, can get tedious to compute. An alternative strategy is to bound the function between two other functions, and hence bound its integral between the integrals of those two functions.

For instance, consider the integral:

$$\int_1^2 \frac{dx}{x}$$

We consider the function $f(x) := x^{-1}$ on the interval $[1, 2]$. Since $x \ge 1$, this function is bounded from above by $g(x) := x^{-1/2}$, and from below by $h(x) := x^{-3/2}$. Thus, the integral of f is bounded between the integrals of g and of h.

An antiderivative for g is $2\sqrt{x}$, and evaluating it between limits gives $2(\sqrt{2} - 1)$, which is slightly less than 0.83. An antiderivative for h is $-2/\sqrt{x}$, and evaluating it between limits gives $2 - \sqrt{2}$, which is slightly greater than 0.58. Thus, the integral of f is somewhere between 0.58 and 0.83. (the actual value is about 0.693, as you will see later). Note how we were able to get a very reasonable estimate without computing an antiderivative or using upper and lower sums.



In fact, upper and lower sums are a special case of this bounding procedure where the two bounding functions that we choose are *piecewise constant functions*.

3.4. **Other inequalities.** Recall the triangle inequality, which states that for any two real numbers x and y, we have:

$$|x + y| \leq |x| + |y|$$

This can be generalized to more than two variables. The general form reads as:

$$|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|$$

Since an integration is an infinite analogue of a sum, the triangle inequality must have an analogue for integration. This reads as follows:

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx$$

Note that we already have a geometric intepretation of both sides. The right side is the total unsigned area between the graph of f and the x-axis from point a to point b. The left side is the magnitude of the signed area from point a to point b. On the left side, we are adding the areas with signs (leading to possible cancellations) and then taking the absolute value in the end. On the right side, we are adding the absolute values to begin with. Thus, there is no scope for cancellation.

4. The beauty of symmetry

4.1. **The role of symmetry.** Before proceeding to the role of symmetry, we first explore how various transformations of the real line affect the value of the integral.

First, how does shifting by h affect integration?

$$\int_a^b f(x+h)\, dx = \int_{a+h}^{b+h} f(x)\, dx$$

This is an example of the chain rule in action, or the u-substitution. What we did is the following: start from the left side, and express $u = x + h$. Then $du/dx = 1$, and $f(x + h)$ becomes $f(u)$. The limits become $a + h$ and $b + h$, so we get:

$$\int_a^b f(x+h)\, dx = \int_{a+h}^{b+h} f(u)\, du$$

Now, however, u is a *dummy variable*, so we can replace this dummy variable by the dummy variable x. The term *dummy variable* is used for a variable that appears as the variable of integration or summation which hence cannot appear anywhere else. The dummy variable is by nature *local* to the integration or summation operation and hence its representing letter can be changed.

Graphically, the area staked out by f between $a + h$ and $b + h$ is the same as the area staked out by $f(x+h)$ between a and b. This is intuitively clear, because the graph of $f(x+h)$ is obtained from the graph of f via shifting left by h.

The other kind of operation that is of interest here is the flip-over, namely, sending x to $-x$. The relevant identity here is:

$$\int_a^b f(x)\,dx = \int_{-b}^{-a} f(-x)\,dx$$

This again follows from a u-substitution.

These two basic ideas give us the interesting results we have on even, odd, and periodic functions:

(1) Suppose f is an odd continuous function on the interval $[-a, a]$. Then, its integral on $[-a, a]$ is 0. Roughly, this is because the integral on the interval $[-a, 0]$ cancels out the integral on the interval $[0, a]$, with each $f(x)$ being canceled by the corresponding $f(-x)$.

(2) More generally, if f is a continuous function on $[p, q]$ with half-turn symmetry about $((p+q)/2, f((p+q)/2))$, then the integral of f on $[p, q]$ is $(q-p)$ times the value $f((p+q)/2)$. Intuitively, this is the average value, and for every deviation above the value, there is a corresponding deviation below the value on the other side.

(3) Suppose f is an even continuous function on the interval $[-a, a]$. Then, its integral on $[-a, a]$ is twice its integral on $[0, a]$. This is because the picture of the function on $[-a, 0]$ is the same as the picture on $[0, a]$ (in the reverse order from left to right, but this does not affect area).

(4) More generally, if f enjoys mirror symmetry about $x = c$, the integral on $[c, c+h]$ equals the integral on $[c-h, c]$.

(5) If f is a periodic function that is continuous and defined for all real numbers, the integral of f over any interval of length equal to the period is the same. If f has a periodic antiderivative, then this integral is zero. If the period is h, and the integral over one period is k, then we can think of k/h as the long-run average value of f. More on this in the next section and in homework problems.

5. Mean-value theorem for integrals

5.1. **Statement of the theorem.** This result states that if f is a continuous function on a closed interval $[a, b]$, then there exists $c \in (a, b)$ such that

$$f(c) = \frac{\int_a^b f(x)\,dx}{b-a}$$

The right side of this expression is the *mean value*, or *average value*, of f on the interval $[a, b]$. Thus, this result simply states that a function attains its mean value somewhere on the interval.

Recall the earlier mean-value theorem:

If F is a function that is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that:

$$F'(c) = \frac{F(b) - F(a)}{b-a}$$

We now explain how the mean-value theorem for integrals follows from the (original) mean-value theorem. The idea is to pick F as an antiderivative for f. Then, $F' = f$, and F satisfies the hypotheses needed to apply the mean-value theorem for derivatives.

The left side of the (original) mean-value theorem is $F'(c)$, which equals $f(c)$. The numerator on the right side is $F(b) - F(a)$, which, by the fundamental theorem of integral calculus, is the same as $\int_a^b f(x)\,dx$. Thus, we get the necessary expression for the mean-value theorem for integrals.

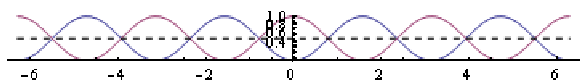
5.2. **Mean value of periodic functions.** For a continuous function defined on all of $\mathbb{R}$, we can define the mean value of the function *over an interval*, but it does not make sense to define an *overall* mean value. For functions that go off to infinity in either direction, the mean value also goes off to infinity as we shift the intervals farther and farther off. On the other hand, for functions that are bounded, there is some hope in talking of a mean value.

One class of functions for which a mean value makes eminent sense are periodic functions. As mentioned earlier, if f is periodic with period h, the integral of f over any interval of length h is a constant. Call this constant k. If F is an antiderivative of f, then F can be expressed as the sum of a periodic and a linear function. The linear part of F has slope k/h. Graphically, F is periodic with shift: the graph of F repeats after a length of h, but is vertically shifted by k.

Thus, there is a strong case to declare that the average value of f is k/h. Note that when f has a periodic antiderivative, then its average value is 0. For instance, sin and cos have average value 0, as we can see from the fact that they are symmetrically distributed above and below the x-axis.

On the other hand, the $\sin^2$ function has positive average value, and its antiderivative has a nontrivial linear component. We'll get back to this function in a short while.

For a periodic function f, it is *not* true that its mean value over *every* interval is k/h. However, any deviation from k/h is due to periodic, or seasonal fluctuation. As far as secular trends go, the mean value is k/h. In particular, if $k \neq 0$ (so that there is a nontrivial linear component) then, in the limit, as interval length becomes large, the mean value approaches k/h, even if the interval length is not a multiple of h. Intuitively, imagine that the period is 1, the average value on an interval of length 1 is k, and we take an interval of length 29417.3. Of this length, if we just took a sub-interval of length 29417, we would get average value k. The remaining interval length of 0.3 can upset things. But the integral on this remaining part will be divided by an interval length of 29417.3, so the deviation it causes will be small. The limit of the average value over an interval, as the interval length goes to ∞, is k/h.



5.3. The $\sin^2$ and $\cos^2$ functions.

A brief note on graphing and integrating the $\sin^2$ and $\cos^2$ functions. Although these functions can be integrated by a method called integration by parts, we will for now use another approach: the double angle formula. This states that:

$$\begin{aligned} \sin^2 x &= (1 - \cos(2x))/2 \\ \cos^2 x &= (1 + \cos(2x))/2 \end{aligned}$$

As a sanity check, note that if we add the right sides, we get 1, as we should.

We can now do graph transformations to plot $\sin^2 x$ and $\cos^2 x$. Note that it is now pictorially clear, even before we bother with actual integration, that both these functions have an average value of 1/2. This stands to reason: the sum of $\sin^2 x$ and $\cos^2 x$ is 1, and they're both the same graph shifted over, so on average, 1/2 should belong to $\sin^2 x$ and the other 1/2 should belong to $\cos^2 x$.

We can also formally integrate these functions:

$$\begin{aligned} \int \sin^2 x\, dx &= (x/2) - (\sin(2x))/4 + C \\ \int \cos^2 x\, dx &= (x/2) + (\sin(2x))/4 + C \end{aligned}$$

We see that the linear part of the antiderivative has slope 1/2, as expected, and the periodic part has periodicity π, again as expected, since $\sin^2$ and $\cos^2$ both have a periodicity of π.

Now, what the discussion about the mean value of periodic functions states is that, over a very long interval, the average value of the $\sin^2$ function is almost 1/2, even if the length of the interval is not a multiples of π.

These formulas for the average value of $\sin^2$ and $\cos^2$ appear in the context of waves. To calculate the energy of a wave involves integrating the square of the wave function over an interval. Since the wave function is of the form $A\sin(mx + \varphi)$, a slight generalization of the above calculations shows that the average energy per unit length of the wave is $A^2/2$. Similarly, if it is a time wave (so $A\sin(kt + \varphi)$) then the average energy per unit time is $A^2/2$. Note that the value of m doesn't affect this energy computation at all, because it is the value of A that affects the average value. (Note: There are different concepts of wave energy, and they usually do depend on the frequency, but the point here is that if the energy is simply defined as the integral of the square of the wave function, then the average value does not depend on the frequency).

6. Fun appendix: statistics application

We will not cover this in class, due to time considerations, but it is suggested you read through this while attempting the advanced homework problems related to this material.

6.1. The Gini coefficient setup.

Recall the setup that we had for the Gini coefficient. We arranged our huge population in increasing order of income. Then, for $x \in [0, 1]$, we defined $f(x)$ as the fraction of the income earned by the bottom x fraction of the population. With reasonable assumptions and using continuous approximations, we obtained that f is continuous and increasing, $f(0) = 0$, $f(1) = 1$, and $f(x) \leq x$ for all $x \in [0, 1]$. These were the observations that were necessary for doing the homework problems.

Another observation that was not necessary for doing the homework problems, but is nonetheless true, is that about the significance of f'. $f'(x)$ measures the fractional contribution of a person at level x (i.e., with x fraction of the population earning less). More precisely, we have:

$$f'(x) = \frac{\text{Income of person at level } x}{\text{Mean income}}$$

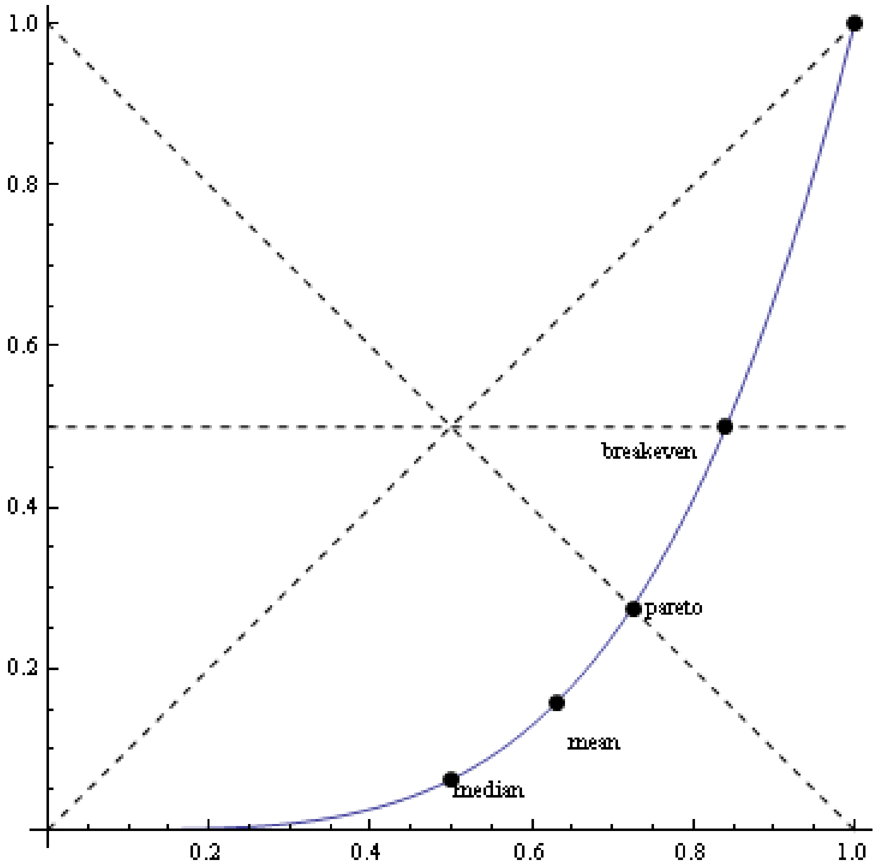
The reason why we need to normalize by mean income is that we have normalized things to $[0, 1]$. Here are some corollaries:

(1) f' is itself increasing, so f is concave up. In other words, people at a higher income level earn more. (In a degenerate case, f' may be constant in an interval, and f linear on that interval. This is when multiple people earn the same income. Unless otherwise stated, we'll assume no degeneracy).
(2) There is a c such that $f'(c) = 1$. This follows from the mean-value theorem. In other words, there is a person who earns the mean income.

6.2. Positions of interest.

What are all the positions x and values $f(x)$ of interest? Here are some of them:

(1) The value x such that the person at level x earns the *mean income*. Mathematically, this means that $f'(x) = 1$. Note that the existence of this value is guaranteed by the mean-value theorem while its uniqueness is guaranteed by the fact that the function is concave up.
(2) The value $1/2$. $f'(1/2)$ is the *median income*. $f(1/2)$ is the fraction of total income earned by the *bottom half* of the population.
(3) The *break-even point*, i.e., the value x such that $f(x) = 1/2$. This is the level x at which the bottom fraction earns the same total income as the remaining top fraction. The break-even point is always bigger than $1/2$ because of the concave up nature of the function. The income earned at this point (given by $f'(x)$) may also be of interest. The existence of a break-even point is guaranteed by the intermediate-value theorem and its uniqueness is guaranteed by the fact that f is increasing.
(4) The *Pareto point*, i.e., the value x such that $f(x) = 1 - x$. This is greater than $1/2$, but less than the break-even point. The income earned at this point may also be of interest. (You proved the existence and uniqueness of this point in your homework).



6.3. **Mean versus median? Mode? Looking at the third derivative.** Is the mean income greater than the median income? Equivalently, is the value x for which $f'(x) = 1$ greater than $1/2$? There is no clear-cut answer. It turns out that the answer depends on whether the distribution of incomes is skewed more toward lower incomes or toward higher incomes.

A third statistical concept that comes up is that of the *mode*. Roughly speaking, the mode is the region where there is maximum clustering of incomes.

We thus want mathematical tools that will help answer the questions: (a) how can we compare mean and median? (b) how can we define mode in this situation?

The answer, interestingly, has something to do with the third derivative.

6.4. **The first, second and third derivatives.** You might remember that, when discussing how to graph functions to understand them better, one useful technique we discussed was to graph the function as well as its first and second derivative (and perhaps higher derivatives as well). Let us put this technique to use here.

Note that the graph of f measures the *cumulative income* earned by certain fractions of the population. This is good for some purposes, but for other purposes, we are interested in individual incomes. Though the graph of f contains this information, it is hidden in that graph. To see the information on individual incomes better, we consider the graph of f'.

As discussed above, the first derivative of f, denoted f', is the ratio of the income of the person at level x to the mean income. We know that f' is a continuous and increasing function on $[0, 1]$. We also know that $f'(0) \geq 0$ and that there is some $c \in (0, 1)$ such that $f'(c) = 1$. Thus, $f'(0) \leq 1$ and $f'(1) \geq 1$. We cannot say anything more conclusive.

Thus, f' is a continuous increasing function on $[0, 1]$ with $0 \leq f'(0) \leq 1$ and $f'(1) \geq 1$. The fact that f' is increasing corresponds to the fact that f is concave up. The value $f'(1/2)$ is the median income, and the point c where $f'(c) = 1$ is the point where the mean income is attained. We can see that the graph of f', subject to the given constraints, could be of many kinds. In particular, the median may occur before the mean or it may occur after the mean.

One advantage of drawing the graph of f' is that, compared to the graph of f, we can focus more in-depth on the way f' increases. We see that f'' measures the rate at which income increases (relative to mean income) as we move from the poorest to the richest. However, we also see that there are many unanswered questions. Where is f' concave up and concave down? Where does it rise most quickly and where does it rise most slowly? We see that the answers to these questions depend on f'''. In the regions where f''' is positive, f' is concave up, which means that the gain in income by moving to the right increases as we move to the right. In the regions where f''' is negative, f' is concave down, which means that the gain in income by moving to the right decreases as we move to the right.

We see that if f''' is positive throughout, that means that the relative gain in income for every slight increase in position goes up as we go from poorer to richer people. This means that the growth of f' is initially sluggish and picks up pace later. Such situations typically correspond to larger values of the mean.

On the other hand, if f''' is negative, that means that the relative gain in income for every slight increase in position goes down as we go from poorer to richer people. In other words, a small step up in the relative ranking means more in income gain terms for poor people than the same small step means for rich people. In this case, the growth of f' is sluggish for rich people and large for poor people. These situations correspond to the mean occurring relatively early.

A final question of interest is about the modal income. What is the income range that most people have? This corresponds to:

- The parts where the graph of f is closest to linear, i.e.,
- The parts where the graph of f' is closest to horizontal, i.e.,
- The parts where the graph of f'' attains its minimum values.
- (Probably) the parts where $f''' = 0$ and $f^{(4))} > 0$.

In other words, the modal segment is the segment where people's income is changing as little as possible with x.

6.5. **The peril of numbers.** Before you entered the world of functions and calculus, the only type of mathematical object you dealt with was a number. But once you entered the world of functions and calculus, you saw yourself dealing regularly with mathematical objects that were more complicated than mere numbers: for instance, sets of numbers, functions, collections of functions, points in the plane (which are ordered pairs of numbers) and so on. Some of these objects are so complicated that it is not possible to describe them using one or two or three numbers.

For instance, we saw that a partition of the interval $[a, b]$ is given by an increasing finite sequence of numbers starting at a and ending at b. Unfortunately, the finite sequence may be arbitrarily large. How do we compare different partitions? We saw two ideas for comparing partitions: (a) The notion of *finer* partition, whereby one refines the other. Unfortunately, given two partitions, it isn't necessary that either one be finer than the other. (b) The notion of the *norm* of a partition, which measures the size of the largest part. We can use the norm to compare two partitions. Unfortunately, a partition with smaller norm may not always behave like a *smaller* partition as far as the upper and lower sums of a particular function are concerned, as you discovered in the midterm.

So, one powerful idea is to use single numbers that measure *size* for complicated objects and reflect some underlying reality of those objects that is empirically useful. The drawback with that idea is that when we look only at that single number, we lose a lot of information about the original object. We may not be able to answer every question that comes up.

The distribution of incomes is another such complicated construct. It is described, as we saw, by this function $f : [0, 1] \to [0, 1]$. But a function cannot be described by a single number. So, instead we ask for single numbers that we can obtain from the function that measure some empirically useful reality about the function. One such number, which tries to measure the *extent of inequality*, is the Gini coefficient. But one problem with the Gini coefficient is that it only measures total inequality, and is not sensitive to inequalities within subpopulations. For instance, if everybody earns roughly the same income and a few people at the top earn a much much larger income, the Gini coefficient is close to 1, even though in some sense there is not much inequality among most people. In other words, the Gini coefficient is sensitive to *huge outliers* in the high-income direction.

That is why it is useful to have a number of different size measures that we can use, and to look at all of them. For instance, the break-even point and Pareto point are useful single numbers that give some intuition about the skew in the distribution of incomes. The median income or the level at which the mean is attained are also useful numbers. When you learn statistics, you will learn many other single numbers that capture useful information about aggregates and distributions. Keep in mind that for any single measure that you choose, there will always be examples of distributions where that measure does not seem to capture what you would like it to intuitively capture.

6.6. **Averages and compositional effects.** As some fun unwinding, here is a trick question. Suppose you have two countries A and B. Is it possible that the mean income in both A and B goes down, but the mean of no *individual* in either country goes down, and in fact, there are individuals whose mean income goes *up*?

Yes, it is possible. Suppose the mean income in country A is 100 money units and the mean income in country B is 400 money units. Imagine that there is a person in country A earning an income of 200 money units who chooses to migrate to country B and gets her income boosted to 300 money units. Assume that nobody else migrates, and nobody else's income changes.

The mean income of country A has gone down, because a person earning above the mean left the country. The mean income of country B has *also* goes down, because it just took in a person earning less than the mean income. The net effect is that both countries see a decline in their mean, but no individual is worse off – and at least one individual is better off! This is just one reason why *group averages and aggregates are not always reflective of individuals*. What we have described here is an example of a *compositional effect* – changes in group compositions affecting averages that reflect the opposite of what is happening at the individual level.

Of course, the group averages might still be useful in their own right, but the statistical error would be to *deduce things about individuals using group averages without taking into account compositional effects and the fluidity of group boundaries*.

Here are some other examples of the same phenomenon:

(1) Inter-sectoral migration: In rapidly industrializing nations such as China, agricultural productivity and industrial productivity are both rising about 5% per year. Yet, overall productivity is rising by something like 8% How is this happening? This is because the industrial sector is much more productive than the agricultural sector. As agricultural productivity increases, less people are needed in agriculture, and so people move from the (comparatively less productive) agricultural sector to the (comparatively more productive) industrial sector. This shifting of people from a less productive to a more productive sector itself causes an increase in productivity independent of the increase in productivity within each sector. Here, agriculture plays the role of the poorer nation A and industry plays the role of the richer nation B.

(2) Inter-level migration in calculus: Imagine that one of you, who is doing badly in the 150s, drops down to the 130s, which are a cakewalk for you. Then, the average mathematical skill of the 150s students increases, the average mathematical skill of the 130s student increases, yet there may probably be a net *decrease* in the overall average mathematical skill of the population, if your mathematical skills decline after you're no longer subjected to the rigors of the 150s.