Executive Summary: Calculus

Paul E. Johnson

August 10, 2015
1. Vocabulary of Relationships.

- Calculus is the study of curves, change, and accumulation.
- Calculus gives us words to describe relationships.
2. Optimization (Find Top or Bottom)

An optimization problem usually has 3 parts.

1. An objective function, say $f$.

2. Although we often write about $f(x)$ as a matter of habit, that's misleading and probably confusing to most new students. Almost all statistical calculations take the data, usually called $x$, as fixed quantities, and instead they adjust coefficients $\beta$ or $\theta$ (or whatever). But I write about $f(x)$ as well, just because it is traditional.

3. Constraints that restrict the extent to which we can adjust the inputs.

Finding the top of a hill is a primary objective in math & engineering after high school.
Choosing the horizontal variable to maximize the vertical.

- We are getting “higher” if the result is increasing
- At the top, we are at a point where we can’t move in either direction without reducing the outcome.
- This is differential calculus!
- An important characteristic of an optimum is that the slope of the objective function is 0 at the optimum point. We are at the top of the hill.
Why is that so Difficult?

- Some functions are easy to optimize & visualize
- Some are difficult, even with a computer
- Some are analytically solvable, so no numerical approximation is needed
- But most of “maximum likelihood” and advanced stats are not “solvable”
3: Understanding Diversity

- This is a probability model for 40 yard dash times.
- Very rare to find a 300 pound man who can run 40 yards in 5 seconds or less.
- How rare is it? Integral Calculus is used to find the area of the shaded area.
Real Numbers, $\mathbb{R}$

- Does it go without saying that we are studying the “real numbers”?
  - Includes all fractions, decimal numbers, integers, and 0
  - Familiar to most school children!

**Definition**

The Real Numbers

Math book will give a definition about a closed group of values, a set where we can add and multiply the members and get back values that are also in the same set. A set of values $(x_1, x_2, x_3,)$

1. $x_1 - x_2$ and $x_1 + x_2$ are in $\mathbb{R}$
2. $x_1 \cdot x_2$ and $x_1/x_2$ are in $\mathbb{R}$
3. There exists a value 1 such that $1 \cdot x_1 = x_1$
4. There exists a value 0 that can nullify any value: $0 \cdot x_1 = 0$. 
Cartesian Plane

- Cartesian plane: All pairs you can get by choosing 2 numbers from $\mathbb{R}$.
- A point is an ordered pair that is represented by a dot in the plane.
- Anticipate confusion because various fields re-name the axes. In political science rational choice models, we often refer to them as $X_1$ and $X_2$ and points are $(x_1, x_2)$, $(y_1, y_2)$ and so forth. This notation may be nicer because, in a 10 dimensional model, it is easier to remember $(x_1, x_2, \ldots, x_{10})$ than to give special letters for every dimension.
Critical Terminology

Critical point value of input at which the output is either at the top of a peak or in the bottom of a valley.

Local vs Global Maximum.
conca
ty: curvature

“conca
down”

“conca
ap”

Important because we use concavity to know if we are at a minimum or a maximum.
Detour: Convex Combinations

- Concave down means “peaked”, generally.
- A more formal definition takes us into a detour on the idea of convex combination.
- Consider a weighting coefficient $\lambda$ that varies from $a$ to $b$
- A convex combination of two points is

$$\lambda a + (1 - \lambda)b$$
A convex combination of two points is

\[ \lambda a + (1 - \lambda) b \]

Please notice the beautiful illustration, which took about 5 hours of fiddling about.
A convex combination of two points is

$$\lambda a + (1 - \lambda)b$$

Please notice the beautiful illustration, which took about 5 hours of fiddling about.
Detour: Convex Combinations

- A convex combination of two points is

\[ \lambda a + (1 - \lambda) b \]

- Please notice the beautiful illustration, which took about 5 hours of fiddling about
Convex Combination in 2 Dimensions

- Points are pairs, like \((x_1, y_1)\), and \((x_2, y_2)\)
- In other things I’ve written, I’ve used notation like \((x_1, x_2)\), \((y_1, y_2)\), \((z_1, z_2)\), so lets be cautious about confusing the presenter.
- Calculate the value of the combination

\[
\lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)
= (\lambda x_1 + (1 - \lambda) x_2, \lambda y_1 + (1 - \lambda) y_2)
\]
Convex Combination in 2 Dimensions

\[ \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) = (\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \]

As \( \lambda \) varies from 1 to 0, the combination “traces” a straight line between the two points.
Convex Combination in 2 Dimensions

\[ \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) \]

\[ = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \]

As \( \lambda \) varies from 1 to 0, the combination “traces” a straight line between the two points.
Convex Combination in 2 Dimensions

\[ \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) \]

\[ = (\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \]

- As \( \lambda \) varies from 1 together 0, the combination “traces” a straight line between the two points
Convexity of Sets Defined Similarly

- A convex set has to be round or egg shaped. It can't have any lacunae or kidney shapes.
- Formally, we say any the convex combination of any pair of points in the set lies entirely within the set.

\[ \text{Any convex combination} \ (\lambda x_i + (1 - \lambda) x_j, \lambda y_i + (1 - \lambda) y_j) \text{ belongs to } X \]
And We Redefine Concave Down

- A function is concave down if we can select any two points on the function’s image and the line connecting them is entirely below the function.

- Phrase that as convex combination of points.
We study relationships.

Inflection Point: where concavity changes.
**Continuity**

**Continuous:** informal def: Can draw without lifting pencil from paper.

Note how the value of the function “hops” at one point.

The open loop indicates that the value of the function \( f(x) \) is undefined.

“Pinholes” like this interfere with analysis, we (usually) assume functions are continuous.
Notation: Open & Closed Intervals

- Closed Interval: \([0, 1]\) all real values between 0 and 1, INCLUDING 0 and 1
- Open Interval: \((0, 1)\) all real values between 0 and 1, NOT INCLUDING 0 and 1
- As long as we know a function’s domain is closed, then we know (for sure!) that the function has a well defined maxima and minima
- On open intervals, minimum may not exist: pick the smallest value inside \((0, 1)\). Undefined!
Outline

1. What is Calculus for?

2. Necessary Terminology

3. Differential Calculus
   - The Big ideas
   - Derivatives I can Remember

4. Optimization
   - First Order Conditions
   - Second-order conditions
   - Multivariate

5. Integration

6. Conclusion
We are Aiming to Understand This

Slope of this tangent line is 0

Objective function $f(x)$

Slope > 0

Slope < 0
Slope of a Straight Line

Comparing two points \((x_1, y_1)\) and \((x_2, y_2)\):

- **difference**
  - horizontally
  \[ \Delta x = x_2 - x_1. \]
  - vertically
  \[ \Delta y = y_2 - y_1. \]

Slope is the ratio of the two changes.

\[
slope = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}
\]

(1)

“the rise over the run.”
Slope is a Local Concept

- A “local” property is a property that is not “true everywhere.” It is not a “global” property.
- The calculation of the slope depends on “where you are,” either on the left or the right of the break point.
Consider the smooth curve.

1. Start at $x_1$
2. Step to $x_2$
3. Stem to $x_3$
Approaching the tangent

Keep making $\Delta x$ smaller and smaller and $\Delta y$ shrinks.

But $\frac{\Delta y}{\Delta x}$ converges to the slope of the line that is tangent to $f(x)$ at $(x, y)$

$x_1, x_2, x_3$ get closer and closer to $x$.

As $x_i$ gets closer to $x$, the slope of the dotted line, $\frac{\Delta y}{\Delta x}$, gets steeper.

“$x$” is a particular point and values $(x_3, y_3)$ $(x_4, y_4)$ $(x_5, y_5)$
The Derivative is the value of $\frac{\Delta y}{\Delta x}$ when $\Delta x$ shrinks to nothing. The dotted line—the tangent line—is “just barely” touching $f(x)$. The derivative is, formally speaking, the slope of the tangent line.

Very Important: This is a smooth, “differentiable” function. It is not jagged, there are no gaps.
Optimization from a Derivative’s point of View

Optimization: Find a point at which the tangent line’s slope is 0

Slope of this tangent line is 0

Slope > 0

Slope < 0

Objective function \( f(x) \)

Slope > 0

Slope < 0
Various Notations for Derivatives

One classic notation is:

\[ \frac{df(x)}{dx} \]  

(2)

or \( y = f(x) \) then:

\[ \frac{dy}{dx} \]  

(3)

More succinct: “f prime of x”, as in:

\[ f'(x) \]  

(4)

My engineering friends like:

\[ Df(x) \]  

(5)
Detour: Limits

Derivatives cannot be defined unless we have the tool known as “the limit”.

- Consider a function $f(x)$
- As $x$ tends to infinity, what does $f(x)$ tend towards?
  It tends toward 4.0, but it never “quite gets there”
- As $x$ tends to 0, $f(x)$ tends to $-\infty$, but it never “gets there”.
- For $0 < x < \infty$, $f(x)$ can be calculated. It is finite.
  - for $x = 50$, $f(x)$ is defined, 3.894

\[ y = 4 - 5.3 \times (1/x) \]
Limit Definition

Notation for limits. When \( x \) is tending to any value \( x_0 \) is

\[
\lim_{x \to x_0} f(x)
\]

The arrow \( \to \) is pronounced “goes to” or “approaches”.

Easy Case: if \( f(x_0) \) exists (is defined, finite) that is the limit.

More Difficult: \( f(x_0) \) may not exist, but a limit may still exist.
As $x$ goes to $\infty$, $1/x$ goes to...

$1/x$ gets smaller as $x$ gets larger, but never reaches 0. However

$$\lim_{x \to \infty} \frac{1}{x} = 0$$
Consider a continuous function $f(x)$.

1. Add “just a bit” to $x$, we arrive at $x + \Delta x$.
2. Resulting change in $y$:

\[ \Delta y = f(x + \Delta x) - f(x) \]  \hspace{1cm} (6)

3. The derivative is defined only if the following limit exists:

\[ \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \]  \hspace{1cm} (7)
In some optimization problems in statistics, we are trying to find the smallest sum of squared errors or to maximize the likelihood of a sample by adjusting 10s or 100s of coefficients.

If we can manipulate the formula so that it is a sum of terms, we can get the separate derivatives and simplify the work.

**Additivity:** The derivative of the sum is the sum of the derivatives.

For functions $f$ and $g$, their combined slope breaks down to the sum of the separate slopes.

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

(8)
Scaling Factors

- Also in statistics, we have data that is rescaled, from Euro to Dollars, Fahrenheit to Celsius, and so forth.
- It is enormously simplifying to know that

\[ \frac{d}{dx} a \cdot f(x) = a \cdot \frac{d}{dx} f(x) = a \cdot f'(x) \]  

(9)
Break Apart Complicated Problems

Complicated expressions can be “broken down”

\[
\frac{d}{dx}\{a \cdot f(x) + b \cdot g(x)\} = a \frac{d}{dx} f(x) + b \frac{d}{dx} g(x)
\]  

(10)
Outline

1. What is Calculus for?
2. Necessary Terminology
3. Differential Calculus
   - The Big ideas
   - Derivatives I can Remember
4. Optimization
   - First Order Conditions
   - Second-order conditions
   - Multivariate
5. Integration
6. Conclusion
Powers of $x$.

1. Slope of line.

   If

   \[ y = b \cdot x \]  \hspace{1cm} (11)

   then (is this too obvious?)

   \[ \frac{dy}{dx} = b \]  \hspace{1cm} (12)

   Digression: Think backwards for a minute. You usually think of $b$ as a constant, but change gears for a minute to notice

   \[ \frac{dy}{db} = x \]
Powers of $x$.

2a. Slope of $x$ squared. If

$y = x^2$ then

$$\frac{dy}{dx} = 2x \quad (13)$$

The slope of $x^2$ is $2x$. 

The tangent at $x = 1$ has slope = 2.

The tangent at $x = -1.25$ has slope = $-2.5$. 

$y = x^2$
Another view of that

\[ y = x^2 \]

- Tangent at \( x = 1 \) has slope = 2
- Tangent at \( x = -1.25 \) has slope = -2.5

The slope of \( x^2 \) is 2x (the red line)
We worked so hard on that one figure...

\[ y = x^2 \]

- Tangent at \( x = 1 \) has slope = 2
- Tangent at \( x = -1.25 \) has slope = -2.5

The slope of \( x^2 \) is 2x (the red line)
Proof of $\frac{d(x^2)}{dx} = 2x$

That is one of the easiest ones to prove and “really believe.” Use the definition:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{2x\Delta x + (\Delta x)^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{(2x + \Delta x) \cdot \Delta x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} 2x + \Delta x$$

$$= 2x$$

Anyway, I believe in that, and take a lot of the rest on faith (insert smiley face here, please).
More Powers of $x$.

2b. Slope of a cube. If

$$y = x^3$$

then

$$\frac{dy}{dx} = 3x^2$$

2c. Slope of $y = x^4$.

$$\frac{dy}{dx} = 4x^3$$

2d. Slope of $y = x^{-1} = 1/x$

$$\frac{dy}{dx} = -1x^{-2}$$

2e. Slope of $\sqrt{x} = x^{1/2}$

$$\frac{dy}{dx} = \frac{1}{2}x^{-1/2}$$
You start to notice a general pattern? Here’s the most important derivative rule:

\[
\frac{d}{dx} x^N = N \cdot x^{N-1}
\]

That is true, whether \( N \) is a whole number or a fraction.
Logarithms

3. The natural logarithm.

\[ \frac{d}{dx} \ln(x) = \frac{1}{x} \quad (21) \]

If you have a log to a different base, say \( \log_{10}(\cdot) \), the derivative involves a “constant of proportionality.”

\[ \frac{d}{dx} \log_b(x) = \frac{1}{\ln(b)} \cdot \frac{1}{x} \quad (22) \]
Recall Euler’s constant, $e$, the base of the natural logarithm.

$$e^x = \exp(x) \quad (23)$$

4. The derivative is:

$$\frac{d}{dx} e^x = \frac{d}{dx} \exp(x) = \exp(x) \quad (24)$$

In other words, you “get the same thing back”.
As in the case of the logarithm, if you are taking powers of some number besides $e$ then a constant of proportionality enters the picture. The book says

$$\frac{d}{dx} b^x = \ln(b) \cdot b^x \quad (25)$$

Note that $\ln(e) = 1$, so this is consistent if $e$ is the base.
Derivative of a product

5. Product rule

\[ \frac{d}{dx} \{ g(x) \cdot h(x) \} = \frac{d}{dx} g(x) \cdot h(x) + g(x) \frac{d}{dx} \cdot h(x) \]  \hspace{1cm} (26)

or, if you like the prime notations,

\[ \frac{d}{dx} \{ g(x) \cdot h(x) \} = g'(x)h(x) + g(x) \cdot h'(x) \]
6. The **chain rule** states that:

$$\frac{d}{dx}\{f(g(x))\} = \frac{df}{dx}\bigg|_{g(x)} \cdot \frac{dg(x)}{dx}$$  \hspace{1cm} (27)

That’s the derivative of $f(x)$ calculated at the location given by the value $g(x)$, multiplied by the derivative of $g(x)$. Confusing enough? Probably. Suppose, for example, you had

$$g(x) = x^2$$

and

$$f(x) = \ln(x)$$

so

$$f(g(x)) = \ln(x^2)$$

$$\frac{d}{dx} f(g(x)) = \frac{1}{x^2} \cdot 2x$$
Outline

1. What is Calculus for?

2. Necessary Terminology

3. Differential Calculus
   - The Big ideas
   - Derivatives I can Remember

4. Optimization
   - First Order Conditions
   - Second-order conditions
   - Multivariate

5. Integration

6. Conclusion
Find where the Derivative equals 0

- In many statistical problems, we can exactly solve to find the optimal parameter estimates.
- This is done by finding a solution for the derivative—the “first order condition”—and doing some followup checking.
Peaks And Valleys

We suppose we are adjusting a parameter, the thing we have been calling $x$ in these notes. We find a point where

$$f'(x) = 0$$  (28)

it means we have found the exact “top of the hill.” (Yeah!)
Or in the “bottom of the bowl.” (Booh!)
Outline

1. What is Calculus for?
2. Necessary Terminology
3. Differential Calculus
   - The Big ideas
   - Derivatives I can Remember
4. Optimization
   - First Order Conditions
   - Second-order conditions
   - Multivariate
5. Integration
6. Conclusion
Second Derivative

Have found a maximum or minimum?

Second Derivative: The derivative of the first derivative. How much is the slope changing?

- Literally, that would be
  \[ \frac{d}{dx} \left\{ \frac{dy}{dx} \right\} \]
  but they usually don’t write it out like that.

  Meaning you start at a critical point \( x \) and go an “itty bitty” amount to the right, how much does the slope change.

  Notation:

  \[ f''(x) = D^2 f(x) = \frac{d^2 y}{dx^2} = \frac{d^2 f(x)}{dx^2} \] (29)

I was trained by people who prefer this kind of notation, \( f''(x) \). But the other is nice too
Telling Your Hill from Your Bowl (Minima and Maxima)

**Minima:** If the slope is getting **bigger** around $x_{min}$, that means the impact of $x$ is “accelerating”. You are in a bowl.

- If $f'(x) = 0$ and $f''(x) > 0$, then $x$ is a local minimum. $f$ is “concave up” at that point.

**Maxima:** If the slope is getting smaller around $x_{max}$, you are heading downward at the top.

- If $f'(x) = 0$ and $f''(x) < 0$, then $x$ is a local maximum. $f$ is “concave down” at $x$. 
Outline

1. What is Calculus for?
2. Necessary Terminology
3. Differential Calculus
   - The Big ideas
   - Derivatives I can Remember
4. Optimization
   - First Order Conditions
   - Second-order conditions
   - Multivariate
5. Integration
6. Conclusion
Suppose we have 3 input variables, $x_1$, $x_2$, and $x_3$.

$$y = f(x_1, x_2, x_3)$$

A **partial derivative** is the change in $f(x_1, x_2, x_3)$ that results when all of the variables are being held constant except one. The most common notations for the partial derivative “with respect to $x_1$” are

$$\frac{\partial y}{\partial x_1}$$

or:

$$f_1(x_1, x_2, x_3)$$

■ Note the subscript indicates the variable under consideration
$m$ First Order Conditions

$f(x_1, x_2, x_3)$ has 3 variables. That means it has 3 first order conditions. We require that ALL OF THE PARTIAL DERIVATIVES equal 0. That is, simultaneously solve

\[
\frac{\partial y}{\partial x_1} = 0 \quad \frac{\partial y}{\partial x_2} = 0 \quad \frac{\partial y}{\partial x_3} = 0
\]

Note, you could as well think of this as a vector of derivatives:

\[
Df = \begin{bmatrix}
\frac{\partial y}{\partial x_1} \\
\frac{\partial y}{\partial x_2} \\
\frac{\partial y}{\partial x_3}
\end{bmatrix} = 0 \tag{31}
\]
One tires quickly of writing down 3 rows of derivatives over and over, so one often just refers to this condition for a maximum or minimum as $Df = 0$.

It is very common in maximum likelihood analysis to call this the “score equation”.

Second order conditions

This is the point at which calculus with many variables turns to hell becomes frustrating. It is easy to calculate a second partial derivative of $f()$ with respect to $x_1$

$$\frac{\partial^2 y}{\partial x_1 \partial x_1} = \frac{\partial^2 y}{\partial^2 x_1}$$

And one can also find the “cross partial” of $\frac{\partial y}{\partial x_1}$ with respect to another variable, say $x_2$.

$$\frac{\partial^2 y}{\partial x_1 \partial x_2}$$

I prefer short hand notation like $f_{11}()$ or $f_{12}()$ for these.
Second order conditions ...

Suppose you begin with the vector of first partials. Differentiate each item by each of the 3 variables. You end up a 3x3 matrix of second partial derivatives like so:

\[
D' = \begin{bmatrix}
  f_{11} & f_{12} & f_{13} \\
  f_{21} & f_{22} & f_{23} \\
  f_{31} & f_{32} & f_{33}
\end{bmatrix} = \begin{bmatrix}
  \frac{\partial^2 y}{\partial x_1 \partial x_1} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \frac{\partial^2 y}{\partial x_1 \partial x_3} \\
  \frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2 \partial x_2} & \frac{\partial^2 y}{\partial x_2 \partial x_3} \\
  \frac{\partial^2 y}{\partial x_3 \partial x_1} & \frac{\partial^2 y}{\partial x_3 \partial x_2} & \frac{\partial^2 y}{\partial x_3 \partial x_3}
\end{bmatrix}
\]  

(32)

This is the so-called **Hessian matrix**.

In practical applications, we find often that

1. Time consuming to calculate the Hessian
2. Calculations are numerically unstable because of digital computing limitations
3. The matrix has some flaw which indicates that we can’t tell if we are at a maximum or not. Software will return an error about the Hessian matrix being “non-positive definite.” That is to say, we can’t say.
Positive Definite Hessians (or the lack thereof)

I have not done a lot of this kind of work, so take this as a “back of the envelope sketch.”

Here’s the idea of a “positive definite” matrix. We are “at” \((x_1, x_2, x_3)\), as indicated by the first order condition.

We’ve got the second derivative matrix

\[
D' = \begin{bmatrix}
f_{11} & f_{12} & f_{13} \\
f_{21} & f_{22} & f_{23} \\
f_{31} & f_{32} & f_{33}
\end{bmatrix}
\]

Take a small perturbation vector \(z\). It is supposed to represent a small change from the location dictated by the first order condition. Calculate the quantity:

\[
[z_1, z_2, z_3] \begin{bmatrix}
f_{11} & f_{12} & f_{13} \\
f_{21} & f_{22} & f_{23} \\
f_{31} & f_{32} & f_{33}
\end{bmatrix} \begin{bmatrix}
z_1 \\
z_2 \\
z_3
\end{bmatrix}
\]

(33)
There's a theorem that says:
If $z'D'z > 0$, then $D'$ is a positive definite matrix. Congratulations, you have a minimum.
If $z'D'z < 0$, then $D'$ is a negative definite matrix. Its a maximum!
This is similar to the univariate case, where $f''(x) < 0$ means you have a maximum.
Horrible Scary S symbol

- The elongated S symbol represents integration.

\[ \int_{a}^{b} f(x) \, dx \]

This means “the total area under the curve \( f(x) \) between \( a \) and \( b \).”

- The symbol \( dx \) represents the “dummy variable of integration.” It is a signal that you are supposed to move along the \( x \) axis when you sum up from \( a \) to \( b \).
Easy: Area under a flat curve

You always drive 50 miles per hour. After 2 hours, you have gone $2 \times 50 = 100$ miles.

Distance traveled is the area under the curve.

$time \times \text{speed (mph)}$. 
Easy: Area under a flat curve

If you always drive exactly 50 miles per hour, the distance traveled, $F(t)$, is easy to calculate.

We know, for example, that at time 10, the distance traveled is 500 miles. And it is also easy to see that the distance traveled between hours 5 and 10 is $F(10) - F(5) = 250$. 
Not So Easy: Variable Speeds

Speed, $f(t)$, is not constant.
Variable Speed

The distance traveled after 10 hours, $F(10)$, is 446.02, and we can see that the trip’s progress from 5 to 10 hours, $446.02 - 249.07 = 196.95$ miles, is quite a bit better than the progress between 10 and 15 hours, $543.61 - 446.02 = 97.59$ miles.
Between Hours 5 and 10

The distance traveled between hours 5 and 10. Using the integral notation,

$$\int_{5}^{10} f(t)dt = 196.95$$  \hspace{1cm} (34)

And that area is equal to $F(10) - F(5)$. 
The functions $f(t)$ and $F(t)$ are linked together. For points $a$ and $b$,

$$\int_{a}^{b} f(t) dt = F(b) - F(a) \quad (35)$$

The accumulator function $F(t)$ is the key.

The Fundamental Theorem of Calculus states that the previous expression is true if $F(t)$ is differentiable and

$$\frac{dF(t)}{dt} = F'(t) = f(t)$$
I’m Unlikely to Prove That

But I can wave my hands wildly

- The difference between the “area under the curve” and the “area of the rectangle below it” are compared.
- Outer edges: $a$ and $b = a + \Delta t$.
- The area of the rectangle $(\Delta t \cdot f(b))$ is easily seen to be smaller than the area under the curve.
Could Adjust the Area Measurement

- Suppose we wanted to leave $a$ and $b$ far apart and use an approximating rectangle.
- We could! Just “top” the rectangle at the correct spot (known as the “Mean Value Theorem”)

**Goldilocks**

- If $f(a)$ gives a rectangle that is “too big,” and
- if $f(b)$ gives a rectangle that is too small,
- $f(c)$ is just right!

\[
\text{Area of rectangle is} \quad \Delta t \cdot f(a)
\]

\[
f(a) \quad f(b) \quad f(c) \quad f(t)
\]

\[
a \quad b = a + \Delta t
\]
Make $\Delta t \to 0$.

- As $\Delta t \to 0$, All three of the rectangles would converge to the same size. See?
  - $b \to a$ and
  - $c \to a$.
  - $\lim_{\Delta t \to 0} f(c) = f(a)$
  - $\lim_{\Delta t \to 0} f(b) = f(a)$.

- Now consider the “just right” area measurement:

  $$F(a + \Delta t) - F(a) = f(c) \times \Delta t$$  \hspace{1cm} (36)

- Divide both sides by $\Delta t$. 

Make $\Delta t \to 0$. ...

\[
\frac{F(a + \Delta t) - F(a)}{\Delta t} = f(c)
\]

(37)

The left hand side is starting to look like a derivative, so take limits.

\[
\lim_{\Delta t \to 0} \frac{F(a + \Delta t) - F(a)}{\Delta t} = \lim_{\Delta t \to 0} f(c)
\]

(38)

\[
\frac{dF(a)}{dt} = f(a)
\]

In words, the derivative of $F(t)$ at point $a$ equals the value of $f(t)$ at $a$.

- If we only had $F(t)$, life would be sweet!
- And we do, sorta...
- $F(t)$ is the function which, when differentiated, would be equal to $f(t)$. The slope of $F(t)$ is $f(t)$. For this reason, $F(t)$ is known as an "anti-derivative" of $f(t)$. 
If we were doing this numerically

We sometimes have no analytical method to calculate $F(x)$, especially in multiple dimensions. So there is a high priority on developing numerical methods that can approximate.
Executive Summary: Calculus
Integration

Finer Rectangles = More Calculation

- Waste more computing time, get a better answer.
- This “brute force” numerical approach only works in low-dimensional problems.
- Advanced optimization jargon in your future
  - Gaussian quadrature: smart ways to choose the rectangle positions
  - Monte Carlo math: estimate area by random sampling (GHK algorithm)
Key ideas: Derivative

- There are many details, but the essence of the differential calculus boils down to this.

**Critical Point**

Derivative: where slope is 0

**Local Maximum**

**Second Order Conditions:** Are we at the top of a hill?
Key ideas: Integrals

■ Possible to calculate (analytically or numerically) the area under a curve.

■ Analytically, the Fundamental Theorem of Calculus gives conditions under which the area can be calculated exactly, without numerical approximation.

■ Many applications in probability have analytical solutions.

■ In practice, the models we want to solve are generally not analytically solvable, so computer approximation is vital.
Where to go from here

- Step number: don't let your skin crawl when somebody says “derivative”, “integral”, or “Calculus”
- If you are going into a field where you will be deriving statistical estimators, or using “cutting edge tools,” NOW is the time to refresh your Calculus skills.