Math for Data Science

Calculus - Integrals

Joanna Bieri DATA100

Important Information

- Email: joanna_bieri@redlands.edu
- Office Hours take place in Duke 209 unless otherwise noted – Office Hours Schedule

Today's Goals:

- Define the integral
- Understand what the integral means
- How is integration used in Data Science

(Review) Derivative

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

The derivative gives us information about the slope of the function (or the slope of a tangent line at a point) and they can be used to find critical points and classify critical points as maximums or minimums. This is especially important in optimization.

Integrals - the big idea

Integrals are the opposite of the derivative. Here is some notation:

$$f(x) = \int F(x)\, dx$$

the s shaped symbol $\int$ tells us to take the integral, the function $F(x)$ is called the integrand, and we are integrating with respect to the variable x based on the fact that we see dx at the end of the integral.

Because the integral is the opposite of the derivative, in the expression above:

$$\frac{df}{dx} = F(x)$$

Example - integral and derivative

Lets say I have a function and I take the derivative

```
def f(x):
    return x**3

x = sp.symbols('x')
y = f(x)
y_p = sp.diff(y,x)
```

$3x^2$

Example - integral and derivative

But then say I wanted to "untake" or "undo" the derivative. I could use the integral to do that

```
# My integrand is the derivative
F = y_p
# This is how I can use sympy to integrate with respect to x
sp.integrate(F,x)
```

x^3

Example - integral and derivative

What this means is that

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

AND

$$\int 3x^2 dx = x^3$$

so if I know something about the rate of change of a variable I can use the integral to go backward to know about he original variable

Applied Example

Let's say I know the growth rate of some bacteria - maybe I got a bunch of data and modeled the growth rate over time.

$$G(t) = 3(t-12)^2 - 10$$

Now I want to use this information to predict the amount of bacteria in a sample at time t. Well I want to go from a function that represents growth (number of cells/hour) to a function that just tells me amount (number of cells). This is opposite of rate of change, this is undoing the rate of change. This is exactly what the integral does!

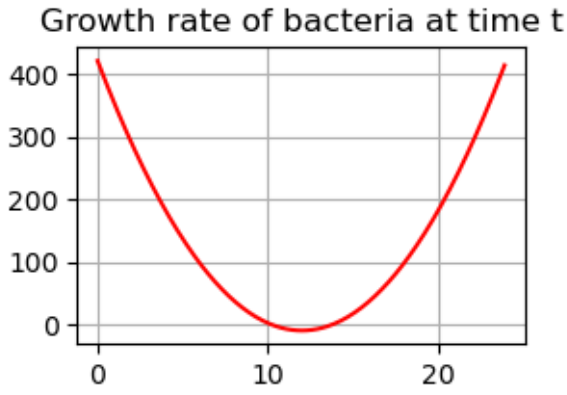
Applied Example

```
def integrand(t):
    return 3*(t-12)**2-10

t = sp.symbols('t')
G = integrand(t)

B = sp.integrate(G,t)
B
```

$t^3 - 36t^2 + 422t$

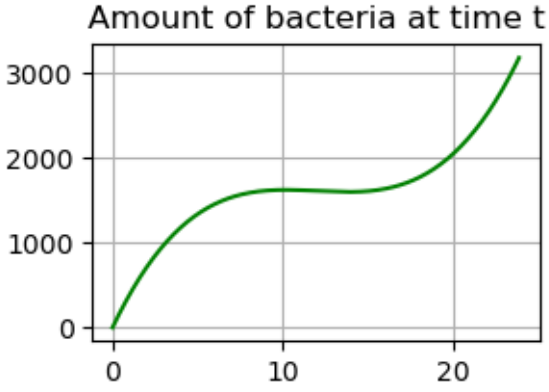
Applied Example

Lets plot these two functions!



Applied Example

Lets plot these two functions!



Interpret these results

The growth rate function (red) shows how fast the bacteria is growing at time t. The bacteria is mostly increasing, except for just a little time in the middle of the day where the derivative is negative. There are two critical points where we have no growth.

The bacteria function (green) which we found by integrating the growth function shows how much bacteria is in the system a time t. We can say exactly how much bacteria we expect to find at time t. We see the function is mostly increasing except for a small range where it slows down.

Interpret these results

INTERESTING RELATIONSHIP BETWEEN GRAPHS It turns out that there is a really interesting relationship between these two graphs! We can see the integral visually on the graph.... let's explore.

Integrals - building up the definition

Let's think for a minute about how we might go between a function that tells me the rate of change and a function that tells me a value. We will build up this idea using some examples.

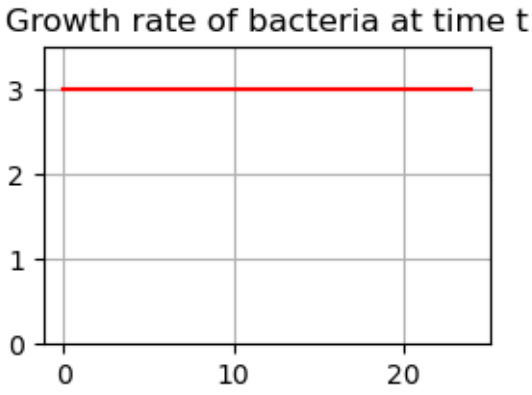
What if we knew that something had a constant rate of change:

$$G(t) = 3$$

bacteria is always growing at a rate of 3 cells per hour.

Example 1 - Constant Rate of Change

Plot this growth rate:



Example 1 - Constant Rate of Change

Now ask a question

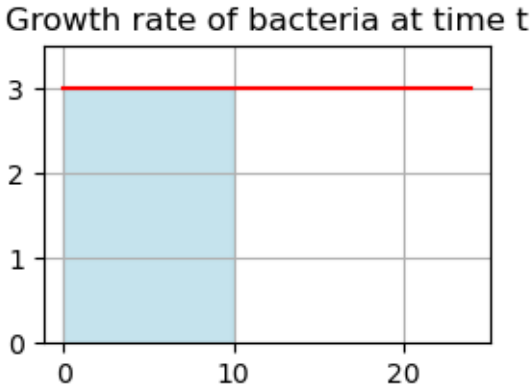
If I knew that at time $t = 0$ I had exactly 5 cells, how many cells would I have when $t = 10$? How would I calculate this? Just using algebra and instincts:

- The rate of change is 3 cells per hour. Every hour we get 3 more cells.
- $t = 10$ represents 10 hours of time elapsing.
- So, my cells would increase by $3 \times 10 = 30$ cells in 10 hours.
- If I started with 5 cells then at hour 10 I would have 35 cells.

But think about this on the graph. What does this multiplication 3×10 look like?

Example 1 - Constant Rate of Change

It's filling in the area under the curve! The total amount of bacteria over that time period is just the area under the curve - BASE times HEIGHT!



Example 1 - Constant Rate of Change

Compare this to the integral.

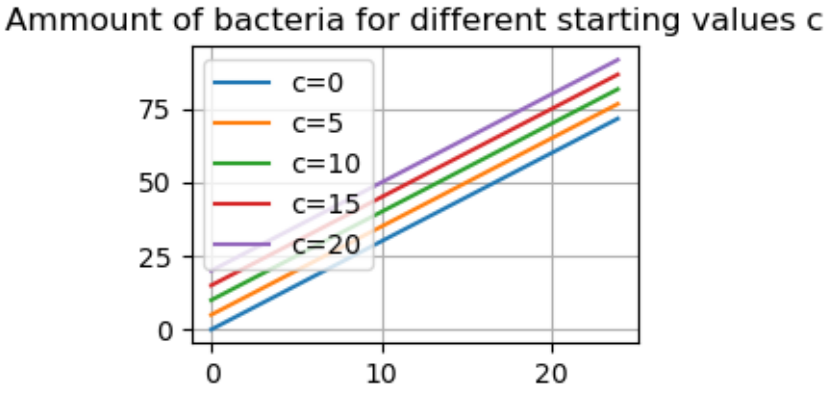
If I integrate to get the solution to this problem, I find:

$$B(t) = \int 3 \, dt = 3t + c$$

Here I added c the constant of integration... just knowing the rate of change is not enough, I need to know where I started to know how much I have. So the amount of bacteria at time t is a straight line function.

Example 1 - Constant Rate of Change

Look at a graph:



Example 1 - Constant Rate of Change

Does this make sense?

- If i started with no bacteria but somehow it was growing at 3 per hour, at hour ten I would have 30 cells.
- If I started with 5, at hour ten I would have 35 cells.
- and so on.

Integration - what do we know

1. The integral of a function is related to the area under the curve.
2. The solution to the integral is a family of functions where the $+c$ represents the starting value.
3. The integral is the opposite of the derivative.

But we didn't need calculus for this problem!

Example 2 - Non-Constant Rate of Change

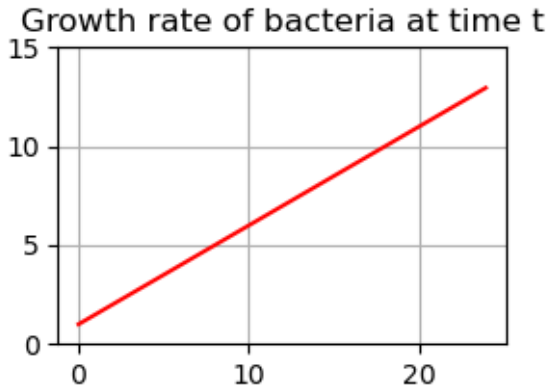
We will redo this example except now we have a non constant rate of change

$$G(t) = \frac{t}{2} + 1$$

bacteria is always growing at an increasing rate. When $t = 0$ there is a growth of 1 cell per hour, when $t = 24$ it is growing at a rate of $24/2 + 1 = 13$ cells per hour.

Example 2 - Non-Constant Rate of Change

Plot this growth rate:



Example 2 - Non-Constant Rate of Change

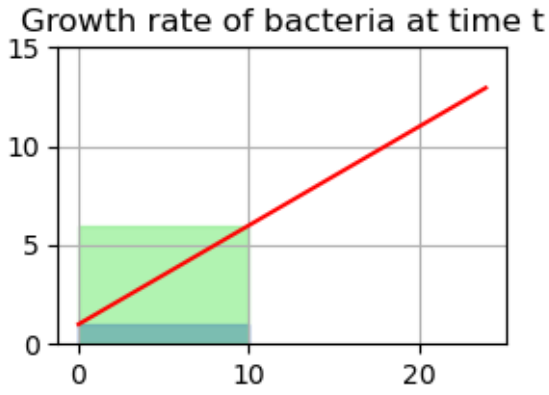
Now ask the same question

If I knew that at time $t = 0$ I had exactly 5 cells, how many cells would I have when $t = 10$? How would I calculate this? Just using algebra and instincts:

- This one is a little harder because the function is changing.
- What rate should I pick?
- If I pick G(0) as the growth rate then it would say I have $1 \times 10 = 10$ cells added or a total of 15 cells - underestimate
- If I pick G(10) it would say I have a growth of 6 cells per hour $6 \times 10 = 60$ cells added or a total of 65 cells- overestimate

Example 2 - Non-Constant Rate of Change

Lets look at a graph - over vs under estimate:

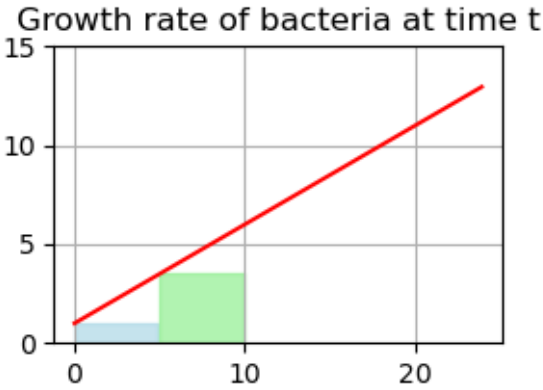


Example 2 - Non-Constant Rate of Change

One of these is way under estimating and the other way over. The problem here is that the growth rate is changing continuously throughout the range. We can't just easily pick a height. How could we do better?

- We could divide the range into two pieces and use two estimates for the rate of change. Lets use $G(0)$ and $G(5)$ as the estimates for the growth and then the base of our rectangle would be 5.

Example 2 - Non-Constant Rate of Change



Example 2 - Non-Constant Rate of Change

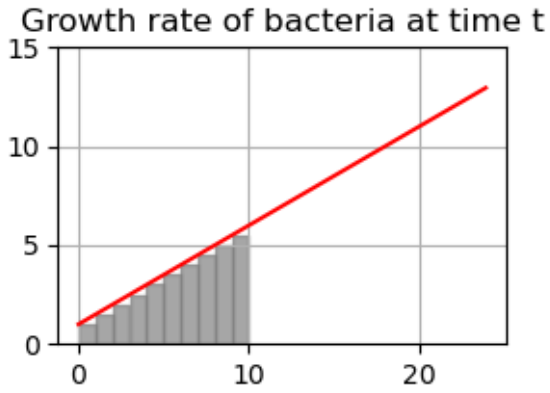
This is probably better, but still an underestimate.

$$B = G(0)(5) + G(5)(5) = 1(5) + 3.5(5) = 22.5$$

So we are estimating that we have $5 + 22.5 = 27.5$ cells at time $t = 10$

Example 2 - Non-Constant Rate of Change

Well if we can divide twice we can divide 10 times, right?




```
Total bacterial at time 10 if we use 10 rectangles: 37.5
```

Example 2 - Non-Constant Rate of Change

Wow, that got a lot better! What is our estimate now?

$$B = G(0) * \Delta t + G(1) * \Delta t + G(2) * \Delta t + \cdots = 32.5 cells$$

where Δt is the width of the base of the rectangle, here $\Delta t = 1$. At $t = 10$ we would have $32.5 + 5 = 37.5$ cells.

You Try - how can we get a better estimate?

- Rerun the code above to get a better estimate. You should only have to change the value of N!

Questions

- As we increase N we are decreasing what?
- What is the general formula for adding up the rectangles (try to use summation notation it's okay if it's not perfect!)
- For a perfect answer we want the number of rectangles to go to what?
- What does this sound like we are doing?

Answers

- As we increase the number of rectangles we are decreasing the width of the rectangles - Δt is getting smaller.

$$\Delta t = (tend - tstart)/N$$

- To add these things up we are doing a sum

$$\sum_{i=0}^{N} G(t_i)\Delta t$$

where we are taking the t value from the left hand side of the function. *We could have just as easily taken this t value from the right side or even the middle somewhere.*

Answers

- To get a perfect answer we would want an infinite number of rectangles with infinitesimal width!
- This is a LIMIT!!!

$$\lim_{N \to \infty}$$

Integral Definition

We define the integral as the limit as N goes to infinity of this sum (Reimann Sum)

$$\int G(t)\,dt = \lim_{N\to\infty} \sum_{i=1}^{N} G(t_i^*)\Delta t$$

Integral Definition

Lets use the integral to answer our question exactly!

Given that the rate of change in the number of cells at time t is

$$G(t) = \frac{t}{2} + 1$$

If I knew that at time $t = 0$ I had exactly 5 cells, how many cells would I have when $t = 10$?

How would I calculate this?

1. Find $B(t)$ using the integral - remembering that it will give us a family of functions (we will need to add $+c$)
2. Plug in our numbers to find the answer

Integral Definition

```
# Enter the growth function symbolically
t = sp.symbols('t')
G = t/2+1

# Let sympy find the integral
sp.integrate(G,t)
```

$$\frac{t^2}{4} + t$$

Integral Definition

This means that the amount of bacteria is given by the family of functions

$$B(t) = \frac{t^2}{4} + t + c$$

we would need to know how much we started with to solve for c. But we do know this! $B(0) = 5$ when $t = 0$ we have 5 cells. PLUG THIS IN

$$B(0) = c = 5$$

so for our particular problem

$$B(t) = \frac{t^2}{4} + t + 5$$

Integral Definition

```
B = t**2/4 + t +5
B
```

$$\frac{t^2}{4} + t + 5$$

We can answer the question by subbing in $t = 10$.

```
B.subs(t,10)
```

$$40$$

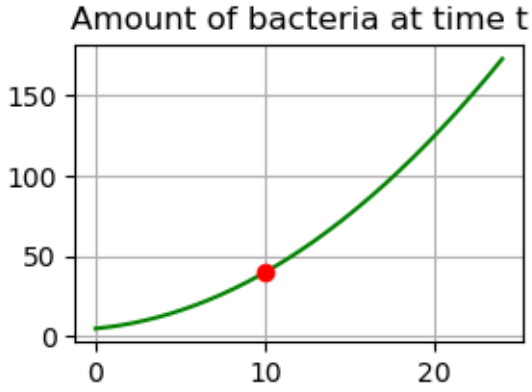
When $t = 10$ I will have 40 cells.

Integral Definition

Because I have a general function $B(t)$ I can actually answer the question for any time! I could easily change the answer if I had a different starting value. Calculus gives me a lot of power and flexibility.

Integral Definition

Lets plot our function (the integration result)



Integration - what do we know

1. The integral of a function is related to the area under the curve.
2. The solution to the integral is a family of functions where the $+c$ represents the starting value, or at some point.
3. The integral is the opposite of the derivative.
4. It is a type of fancy addition - we are adding up the area under the curve of a changing function.

Example 3 - Non-Constant Rate of Change

Let's take this a step further and use our original function

Let's say I know the growth rate of some bacteria - maybe I got a bunch of data and modeled the growth rate over time.

$$G(t) = 3(t-12)^2 - 10$$

Now I want to use this information to predict the amount of bacteria in a sample at time $t = 10$ and I know that when $t = 0$ I had 5 cells.

Example 3 - Non-Constant Rate of Change

Here are my steps:

1. Enter the growth function symbolically
2. Use sympy to take the integral
3. Find the constant of integration c using the information given.
4. Construct the solution for your problem (particular soln)
5. Answer the question.

Example 3 - Non-Constant Rate of Change

We can do this all symbolically - using Python

```
# 1. Enter the growth function
t = sp.symbols('t')
def growth(t):
    return 3*(t-12)**2-10

G = growth(t)
G
```

$$3\left(t-12\right)^{2}-10$$

Example 3 - Non-Constant Rate of Change

```
# 2. Find the integral
B = sp.integrate(G,t)
B
```

$t^3 - 36t^2 + 422t$

Example 3 - Non-Constant Rate of Change

$$B(t) = t^3 - 36t^2 + 422t + c$$

after I add my constant of integration. I plug in $B(0) = 5$

$$B(0) = 0^3 - 36 * 0^2 + 422 * 0 + c$$

$$5 = (0)^3 - 36(0)^2 + 422(0) + c$$

so

$$c = 5 - [(0)^3 - 36(0)^2 + 422(0)]$$

Example 3 - Non-Constant Rate of Change

in Sympy I could write this as

```
# 3. Find the constant
tknown = 0
Bknown = 5

c = Bknown - B.subs(t,tknown)
c
```

5

Example 3 - Non-Constant Rate of Change

```
# 4. Enter my B function with the added c
B = B + c
B
```

$$t^3 - 36t^2 + 422t + 5$$

```
# 5. Answer the question
B.subs(t,10)
```

$$1625$$

Example 3 - Non-Constant Rate of Change

Interpret Result

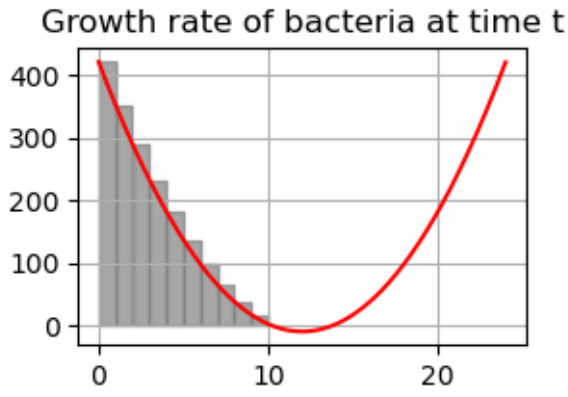
If my bacteria is growing at a rate given by $G(t)$ and I start with 5 cells at time $t = 0$ then I will have 1625 cells ten hours later.

Visualization - extra

You don't have to visualize the integral every time but it is important that you remember that it represents the area under the curve! We are adding up the growth function for every single teeny tiny narrow rectangle!

Visualization - extra

If you play with the code below you should see that as you increase N you get closer and closer to the real value of the integral!




```
Total bacterial at time 10 if we use 10 rectangles: 1840.0
```

You try

Solve the following problems using the integral. For each you should:

1. Enter the growth function symbolically
2. Use sympy to take the integral
3. Find the constant of integration c using the information given.
4. Construct the solution for your problem (particular soln)
5. Answer the question.

Thanks to google gemini for the inspiration for these problems

You try

A city's water reservoir is experiencing varying inflow rates due to seasonal changes. The rate of water inflow, measured in cubic meters per day, is given by the function:

$$I(t) = 1500 + 300 * sin(\pi * t/180)$$

where I is measured in cubic meters per day and t is measured in days.

- If the reservoir contained $50,000$ cubic meters of water at $t = 0$, determine the function representing the total volume of water in the reservoir at any time t.
- Calculate the total volume of water that entered the reservoir between day 0 and day 90. Subtract the amount at $t = 0$ from the amount at $t = 90$.
- How much water should the city have on day 180?

You try

A chemical reaction is producing heat at a rate that changes over time. The rate of temperature increase, measured in degrees Celsius per minute, is given by the function:

$$T(t) == 0.8e^{-0.05t} + 0.1$$

- If the initial temperature at $t = 0$ was 25 degrees Celsius, determine the function representing the temperature at any time t.
- What is the temperature of the reaction after 30 minutes?

You try

A factory's production rate of a certain product is changing over time. The rate of production, measured in units per hour, is given by the function:

$$P(t) = 0.1t^2 - t + 10$$

where t is the time in hours between 0 and 24.

- If the factory had produced 0 units before time $t = 0$, determine the function representing the total number of units produced at any time t.
- The cost of running the factory is given by the function

  $$C(t) = 0.5t^3$$

  and if we imagine that we can sell these products instantaneously for 2 dollars a piece then the profit is given by

Integration and Data Science

In data science we don't calculate integrals very often. What is far more important is the idea of what the integral stands for:

1. It finds the area under a curve
   - probability - total area under curve is one
2. It is a way to do fancy addition of a continuously changing function
   - signal processing
   - probability
   - certain cost functions rely on the idea of integration.
3. It is the opposite of the derivative
   - If we know something about the rate of change in our data we can know about the original data.