Mathematics Review

Brian Vegetabile

2016 Statistics Bootcamp
Department of Statistics
University of California, Irvine

September 6th, 2016

Goals for Lecture

- Familiarize everyone with functions
- Review differentiation and integration
- Review basic optimization of functions
- References: Abbott [2001], Casella and Berger [2002]

Note - that this is a gentle reminder of the topics that you should have learned in your undergraduate mathematics courses. It does not cover everything.

Defining Functions

- In the first lecture, we used functions without properly defining them.
- Here we dive deeper into some of the mathematical definitions needed for statistics

Definition (Function)

Given two sets A and B, a function from A to B is a rule or mapping that takes element $x \in A$ and associates it with a single element of B. We write

$$f : A \longrightarrow B$$

Given an element $x \in A$, the expression $f(x)$ is used to represent the element of B associated with x by the function f.

Properties of Functions

- We now reintroduce some familiar definitions that are useful throughout statistics.
- Consider a function $f : A \to B$.

Definition (Domain of a Function)

The set A is called the domain of f.

Definition (Range of a Function)

The range is the subset of B given by

$$\{y \in B : y = f(x) \text{ for some } x \in A\}$$

Examples of Parameterized Functions I

- Consider the following parameterized function $f : \{0, 1\} \to (0, 1)$

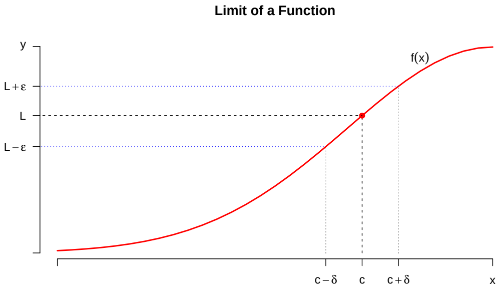
$$f(x|p) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

- The domain is the set $A = \{0, 1\}$ and the range is $\{p, 1 - p\}$.
- *Note:* When we parameterize a function, we imply that the function is dependent on values of the parameter, which we often "know" beforehand. When we know the value of the parameter θ, then we say that the value is "given" and use the notation $f(\cdot|\theta)$.

Defining the Limit of a Function

Definition

Let $f : A \to \mathcal{R}$, and let c be a limit point of the domain A. We say that $\lim_{x \to c} f(x) = L$ provided that for all $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < |x - c| < \delta$ (and $x \in A$) it follows that $|f(x) - L| < \epsilon$.



Defining The Exponential Function

- We can use limits to define certain functions, or show convergence of functions throughout statistics.
- One useful relationship is the definition of the exponential function

Definition (exponential function)

$$\exp(a) = e^a = \lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n$$

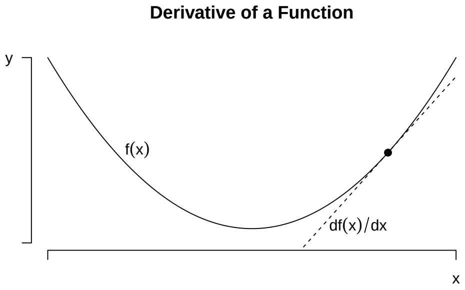
- *(Note: this limit may serve you well through graduate statistics...)*

Overview of Differentiation

- Derivatives play a major role throughout statistics.
- A major area that they become useful is for maximum likelihood estimation.
- Most of frequentist statistics relies on taking derivatives (Score Functions, Information Matrix, etc.)

Intuition of a Derivative

- The geometric interpretation of the derivative of a function f at a point x is that it is the slope of the function at the point x.
- If $f(x)$ is the value of the function evaluated at point x, then the slope at the point x is denoted $f'(x)$.



Defining the Derivative

- We now provide the formal definition of a derivative

Definition (Derivative)

Let $g : A \to \mathcal{R}$ be a function defined on an interval A. Given a point $c \in A$, the derivative of g at c is defined by

$$g'(c) = \lim_{x \to c} \frac{g(x) - g(c)}{x - c}$$

provided that the limit exists. If g' exists for all points $c \in A$, we say that g is differentiable on A.

Derivative Rules I

- Note that these "rules" are not strictly correct in mathematical analysis sense and more follow what was taught in your undergraduate calculus curriculum.
- It is worth reviewing the rigorous mathematical definitions and theorems for these rules.

Derivative Rules II

i) Derivative of a constant function
If $g(x) = c$, that is the function takes the same value for all points $x \in A$, then

$$g'(x) = \frac{d}{dx}c = 0$$

ii) Derivative of a 'Power' function
If $g(x) = x^n$, where n is any real number, then

$$g'(x) = \frac{d}{dx}x^n = nx^{n-1}$$

Derivative Rules III

iii) Derivative of sums of functions
If f and g are both differentiable functions then,

$$\begin{aligned}(f+g)'(x) &= \frac{d}{dx}\left(f(x)+g(x)\right)\\ &= \frac{d}{dx}f(x)+\frac{d}{dx}g(x)\\ &= f'(x)+g'(x)\end{aligned}$$

iv) Derivative when there is a constant multiple of the function
If g is a function and c is some constant real value, then

$$(cg)'(x) = \frac{d}{dx}cg(x) = c\left(\frac{d}{dx}g(x)\right) = cg'(x)$$

for all $c \in \mathcal{R}$.

Derivative Rules IV

v) Multiplication rule for differentiation
Consider two functions g and f, then

$$(fg)'(x) = \frac{d}{dx}\left(f(x) \times g(x)\right) = f(x)g'(x) + g(x)f'(x)$$

vi) Quotient Rule for differentiation Consider two functions g and f, then

$$\left(\frac{f}{g}\right)'(x) = \frac{d}{dx}\frac{f(x)}{g(x)} = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

provided that $g(x) \neq 0$ for all $x \in A$

Derivative Rules V

vii) Chain rule for differentiation
If g is differentiable at x and f is differentiable at $g(x)$, then the composite function $(f \circ g)(x) = f(g(x))$ is differentiable at x and is given by the product

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

In alternate notation, if $y = f(\theta)$ and $\theta = g(x)$ are both differentiable functions, then

$$\frac{dy}{dx} = \frac{dy}{d\theta}\frac{d\theta}{dx}$$

(The chain rule is incredibly useful when attempting to find the maximum likelihood)

Derivative Rules VI

viii) Derivatives of exponentials and logarithmic functions
Consider the function $f(x) = \log(x)$ (here we assume that $\log$ is the natural logarithm), then the derivative is

$$f'(x) = \frac{d}{dx}\log(x) = \frac{1}{x}$$

Consider the function $g(x) = e^x$, then the derivative of the function is

$$f'(x) = \frac{d}{dx}e^x = e^x$$

Example - Chain Rule in Action I

Consider the following parameterized function

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

and we would like to find the derivative with respect to the variable x.

Example - Chain Rule in Action II

We could think of the function in the following way.

$$\begin{aligned} f(g(x)) &= constant \times \exp\{g(x)\} \\ g(x) &= -\frac{(x-\mu)^2}{2\sigma^2} \end{aligned}$$

Example - Chain Rule in Action III

From the above rules, we have that

$$\begin{aligned} f'(g(x)) &= constant \times \exp\{g(x)\} \\ g'(x) &= -\frac{(x-\mu)}{\sigma^2} \end{aligned}$$

Thus by the chain rule,

$$\begin{aligned} f'(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \times \frac{-(x-\mu)}{\sigma^2} \\ &= -\frac{(x-\mu)}{\sigma^2\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \end{aligned}$$

Comparing the Product & Quotient Rules I

- Consider the function

$$f(x) = xe^{-x} = \frac{x}{e^x}$$

we can evaluate its derivative using the quotient or the product rules

Comparing the Product & Quotient Rules II

- Via the product rule

$$h'(x) = f(x)g(x) \implies h'(x) = f(x)g'(x) + f'(x)g(x)$$

therefore,

$$f'(x) = -xe^{-x} + e^{-x} = e^{-x}(1-x)$$

- By the quotient rule,

$$h'(x) = \frac{f(x)}{g(x)} \implies h'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g'(x))^2}$$

therefore,

$$f'(x) = \frac{e^x - xe^x}{(e^x)^2} = \frac{(1-x)}{e^x} = e^{-x}(1-x)$$

Multiple Differentiation Example I

- As the year progresses we'll be interested in finding derivatives of functions with many parameters
- Specifically, as will be defined later this year the score function and the Fisher Information matrix are defined in terms of derivatives.
- We'll review the notation for derivatives of multiple variables.

Multiple Differentiation Example II

- From Ferguson's 'A Course in Large Sample Theory'
- Consider $f : \mathcal{R}^d \to \mathcal{R}$, the derivative of f is the row vector

$$f'(\mathbf{x}) = \frac{d}{d\mathbf{x}} f(\mathbf{x}) = \left(\frac{d}{dx_1} f(\mathbf{x}), \frac{d}{dx_2} f(\mathbf{x}), \ldots, \frac{d}{dx_d} f(\mathbf{x}) \right)$$

- The second derivative of $f : \mathcal{R}^d \to \mathcal{R}$ is

$$f''(\mathbf{x}) = \frac{d}{d\mathbf{x}} f('\mathbf{x})^{\mathbf{T}} = \begin{pmatrix} \frac{d^2}{dx_1 dx_1} f(\mathbf{x}) & \frac{d^2}{dx_1 dx_2} f(\mathbf{x}) & \ldots & \frac{d^2}{dx_1 dx_d} f(\mathbf{x}) \\ \vdots & \ddots & & \vdots \\ \frac{d^2}{dx_d dx_1} f(\mathbf{x}) & \frac{d^2}{dx_d dx_2} f(\mathbf{x}) & \ldots & \frac{d^2}{dx_d dx_d} f(\mathbf{x}) \end{pmatrix}$$

Multiple Differentiation Example III

- Consider $f(x, y) = x^2 + y^2 - 2xy$, the derivative of the function will be the vector

$$\begin{aligned} f'(x, y) &= \left(\frac{d}{dx} f(x, y), \frac{d}{dy} f(x, y) \right) \\ &= (2x - 2y, 2y - 2x) \end{aligned}$$

- The second derivative will be the matrix

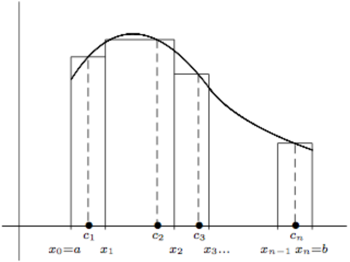
$$f''(x, y) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

Why Integration?

- Integration additionally plays an important role in statistics.
- While, the derivative is the slope of a function a point and often represents the rate of change of the function. The integral represents the area under the curve
- Properties of the random variables are defined based on integration, such as the cumulative distribution function
- Additionally, most of Bayesian statistics relies on integration to find estimates for parameters.

Intuiition of the Integral

- Integration and 'area' go hand in hand. The integral of a function allows you to find the total area under the curve.



- You'll remember from your calculus class the idea of Riemann sums.
- Those of you in the Ph.D. program will further these ideas with measure theoretic integration.

Fundamental Theorem of Calculus

- We'll state the fundamental theorem of calculus as a reminder and then provide some common integration strategies that you learned in your undergraduate work.
- As a reminder the Fundamental Theorem of Calculus allows us to relate integration and differentiation.
- Again this isn't everything you'll need to get through the year, just a reminder of the mathematical tricks that you've learned along the way.

Fundamental Theorem of Calculus

Theorem (Fundamental Theorem of Calculus)

i) *If* $f : [a, b] \to \mathcal{R}$ *is integrable, and* $F : [a, b] \to \mathcal{R}$ *satisfies* $F'(x) = f(x)$ *for all* $x \in [a, b]$, *then*

$$\int_a^b f = F(b) - F(a)$$

ii) *Let* $g : [a, b] \to \mathcal{R}$ *be integrable and define*

$$G(x) = \int_a^x g$$

for all $x \in [a, b]$. *Then* G *is continuous on* $[a, b]$. *If* g *is continuous at some point* $c \in [a, b]$ *then* G *is differentiable at* c *and* $G'(c) = g(c)$.

Common Integration Rules I

Substitution Rule

If $u = g(x)$ is a differentiable function whose range is an interval I and f is continuous on I, then

$$\int f(g(x))g'(x)dx = \int f(u)du$$

For definite integrals, if g' is continuous on $[a, b]$ and f is continuous on the range of $u = g(x)$, then

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

Common Integration Rules II

Integration by Parts

- Based upon the product rule of differentiation we can create a similar rule for integration, namely integration by parts.
- This rule allows us to calculate the integral of the product of two functions and is particularly useful when one of the functions is easy to integrate.

By the product rule of differentiation,

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

Which in term implies that

$$\int \left[\left[f(x)g'(x) + g(x)f'(x)\right] dx = f(x)g(x)\right.$$

Common Integration Rules III

Rearranging

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx$$

For definitely integrals this becomes,

$$\int_a^b f(x)g'(x)dx = f(x)g(x)\Big]_a^b - \int_a^b g(x)f'(x)dx$$

Example - Integration by Parts I

- Consider the function $h(x) = xe^x$ which appears to be unruly to integrate at first glance.
- Using integration by parts, we can actually solve this: Consider,

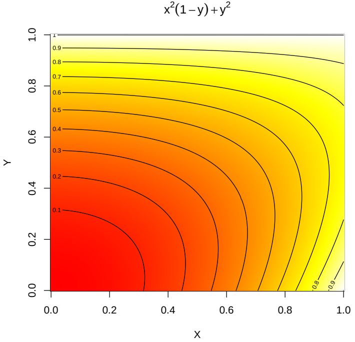
$$h(x) = f(x)g'(x) = xe^x$$
$$f(x) = x \qquad g'(x) = e^x$$

Recall that $\int e^x dx = e^x$ and $f'(x) = 1$, thus using integration by parts

$$\begin{aligned} \int xe^x dx &= xe^x - \int e^x dx \\ &= xe^x - e^x + C \\ &= (x-1)e^x + C \end{aligned}$$

Multiple Integration Example I

- Consider the function $f(x, y) = x^2(1 - y) + y^2$ for $x \in (0, 1)$ and $y \in (0, 1)$.
- We want to find the area under the surface created by the function.



Multiple Integration Example II

$$
\begin{aligned}
& \int_0^1 \int_0^1 x^2(1-y) + y^2 dy dx \\
= & \int_0^1 x^2(y - \frac{y^2}{2}) + \frac{y^3}{3}\Big|_0^1 dx \\
= & \int_0^1 \frac{x^2}{2} + \frac{1}{3} dx \\
= & \frac{x^3}{6} + \frac{1}{3}x\Big|_0^1 = \frac{1}{2}
\end{aligned}
$$

Why Optimization?

- We now talk about finding extreme values of a function.
- This will become very important for maximum likelihood, where want to find the value that maximizes our likelihood function.
- Additionally, this can be used to find the 'mode' (or maximum) of probability density functions.

Defining Extreme Values

Definition (maximum value)

A function f has an absolute maximum (or global maximum) at a point c if $f(c) \geq f(x)$ for all x in the domain of f. The value of $f(c)$ is called the maximum value.

Definition (minimum value)

A function f has an absolute minimum (or global minimum) at a point c if $f(c) \leq f(x)$ for all x in the domain of f. The value of $f(c)$ is called the minimum value.

Definition (extreme values)

The minimum and maximum values are the extreme values of the function.

Interior Extremum Theorem

Theorem (Interior Extremum Theorem)

Let f be differentiable on an open interval (a, b). If f attains a maximum value at some point $c \in (a, b)$. $(i.e. f(c) \geq f(x)$ for all $x \in (a, b))$, then $f'(c) = 0$. The same is true if $f(c)$ is a minimum value.

- The Interior Extremum Theorem is one of the fundamental theorems that enable the use of the derivative as a tool for solving applied optimization problems [Abbott, 2001].
- Always remember to consider the domain of the function as well when finding extreme values.

Example - Optimizing a Function I

Consider the following parameterized function

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Example - Optimizing a Function II

We saw earlier that taking the derivative of this function gave us the following derivative,

$$\begin{aligned} f'(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\} \times \frac{-(x-\mu)}{\sigma^2} \\ &= -\frac{(x-\mu)}{\sigma^2\sqrt{2\pi\sigma^2}} \exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\} \end{aligned}$$

This looks like a mess to find $f'(x|\mu,\sigma^2) = 0$.

Example - Optimizing a Function III

- Alternatively, we can use a strategy which will help us to find derivatives of functions.
- We appeal to monotonic transformations of functions which allow us to preserve the order of the set.
- That is if $g(\cdot)$ is a monotonic function, then if $f(c)$ is an extreme value of the function f, it is also the extreme value off the function g.
- The main monotonic transformation that we will use is a logarithmic transformation.

Example - Optimizing a Function IV

Applying this

$$\begin{aligned} f(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\ \log f(x|\mu,\sigma^2) &= -\log\left(\sqrt{2\pi\sigma^2}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\ \frac{d}{dx}\log f(x|\mu,\sigma^2) &= -\frac{x-\mu}{\sigma^2} \end{aligned}$$

If we set this equal to zero, it will be MUCH easier to solve than the previous function. Thus

$$\begin{aligned} -\frac{x-\mu}{\sigma^2} &= 0 \\ x-\mu &= 0 \\ x_{extreme} &= \mu \end{aligned}$$

Example - Optimizing a Function V

Now we must prove that it is a maximum or minimum! We can use the second derivative.

$$
\begin{aligned}
\log f(x|\mu,\sigma^2) &= -\log\left(\sqrt{2\pi\sigma^2}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\
\frac{d}{dx}\log f(x|\mu,\sigma^2) &= -\frac{x-\mu}{\sigma^2} \\
\frac{d^2}{dx^2}\log f(x|\mu,\sigma^2) &= -\frac{1}{\sigma^2}
\end{aligned}
$$

We see that this is **negative** and therefore a maximum of the function. Thus,

$$x_{max} = \mu$$

Optimization in Action: Newton-Raphson I

- Consider a function f where for some x, $f(x) = 0$, such an x is a root of the function f.
- We can use numerical methods to find such a point, especially if finding such points are difficult.
- Consider the Taylor series expansion of a function around the point x_0

$$
\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \\
&= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots
\end{aligned}
$$

Optimization in Action: Newton-Raphson II

- This implies that if x and x_0 are close then

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

- Further if $f'(x_0)$ is invertible, with $f(x) = 0$ for finding the root, we have that

$$0 \approx f(x_0) + f'(x_0)(x - x_0)$$

- Therefore

$$\begin{aligned} -f(x_0) &= f'(x_0)(x - x_0) \\ -(f'(x_0))^{-1} f(x_0) &= x - x_0 \\ x &= x_0 - (f'(x_0))^{-1} f(x_0) \end{aligned}$$

Optimization in Action: Newton-Raphson III

- Now since

$$x = x_0 - (f'(x_0))^{-1} f(x_0)$$

we can use this to create an algorithm to find roots of the function

- Therefore

$$x_{n+1} = x_n - (f'(x_n))^{-1} f(x_n)$$

- Now since a function will have an extremum when $f'(x) = 0$, we can use this algorithm to find extreme values of functions!
- We'll revisit this later...

References

Stephen Abbott. *Understanding analysis*. Springer, 2001.

George Casella and Roger L Berger. *Statistical inference*. Second edition, 2002.

Background Material

Defining a Sequence

Definition (sequence)

A sequence is a function whose domain is the natural numbers $\mathbf{N} = \{1, 2, 3, 4, 5, \ldots\}$.

- Given a function $f : \mathbf{N} \to \mathcal{R}$, $f(n)$ is the n^{th} term of the list. We note that $\mathcal{R}$ is the real numbers.

Examples of Sequences

- Below are some examples of sequences

i) $\left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots\right)$

ii) $\left(\frac{n+1}{n}\right)_{n=1}^{\infty} = \left(\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots\right)$

iii) (a_n), where $a_n = 2^n$ for $n \in \mathbf{N}$

iv) (x_n), where $x_1 = 3$ and $x_{n+1} = \frac{x_n+1}{2}$

Convergence of a Sequence

Definition (Convergence of a Sequence)

A sequence (x_n) converges to a number c if for every positive number ϵ, there exists an $N \in \mathbf{N}$ such that whenever $n \geq N$ it follows that $|x_n - c| < \epsilon$

- The general idea is that a "challenger" chooses some value ϵ, then you must produce a value N such that for all numbers $n \geq N$ it follows that $|x_n - c| < \epsilon$

Definition (Divergent Sequence)

A sequence that does not converge is said to diverge.

Example of Showing Convergence I

Claim:

$$\lim_{n\to\infty}\left(\frac{n+1}{n}\right)=1$$

Our sequence here is $\left(\frac{n+1}{n}\right)_{n=1}^{\infty} = \left(\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots\right)$, where it appears that for large numbers the numerator and the denominator will be close enough that the value is almost 1. Our goal is find N so that for any ϵ that someone chooses we can satisfy the requirements of the definition for convergence.

Example of Showing Convergence II

Thus we want

$$\begin{aligned}
\left|\frac{n+1}{n} - 1\right| &< \epsilon \\
\left|\frac{n+1}{n} - \frac{n}{n}\right| &< \epsilon \\
\left|\frac{n+1-n}{n}\right| &< \epsilon \\
\left|\frac{1}{n}\right| &< \epsilon
\end{aligned}$$

Now since n is always positive it follows that $1/n < \epsilon$ and $n > 1/\epsilon$. Therefore we can pick N larger than $1/\epsilon$ and then every $n \geq N$ will satisfy the definition. See Abbott page 42 for a formal proof.

Infinite Series I

Definition (Infinite Series)

Let (b_n) be a sequence. An infinite series is a formal expression of the form

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + b_3 + \dots$$

Definition (Partial Sums)

Define the sequence of partials sums (s_m) as

$$s_m = b_1 + b_2 + \cdots + b_{m-1} + b_m$$

Infinite Series II

Definition (Convergence of Series)

We say that the series $\sum_{n=1}^{\infty} b_n$ converges to B if the sequence (s_m) converges to B. We write $\sum_{n=1}^{\infty} b_n = B$

Partial Sum of Integers I

- Let $(a_n) = n$ with $n \in \mathbf{N}$. Then the sequence of partial sums b_m is

$$b_m = \sum_{n=1}^{m} n = 1 + 2 + \cdots + (m-1) + m = \frac{m(m+1)}{2}$$

Note that this does not converge.

- To show this consider that

$$\begin{aligned} b_m &= 1 + 2 + \cdots + (m-1) + m \\ &= m + (m-1) + \cdots + 2 + 1 \end{aligned}$$

Partial Sum of Integers II

- Therefore if we consider $2b_m$, adding the top line and the bottom line of the above we get

$$2b_m = (m+1) + (m+1) + \cdots + (m+1) = m(m+1)$$

- Thus,

$$b_m = \frac{m(m+1)}{2}$$

- Additionally, the partial sums of squared integers is

$$\sum_{n=1}^{m} n^2 = \frac{m(m+1)(2m+1)}{6}$$

we leave this for you to figure out.

Harmonic Series

- Define the harmonic series as $\sum_{n=1}^{\infty} \frac{1}{n}$ with $n \in \mathbf{N}$.
- The sequence of partial sums b_m for the harmonic series is

$$b_m = \sum_{n=1}^{m} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}$$

- Note that this does not converge nor does it have a closed form solution.

Geometric Series

- The geometric series can be written

$$\sum_{n=1}^{\infty} ar^{n-1}$$

- The partial sum of the geometric series is

$$\sum_{n=1}^{m} ar^{n-1} = \frac{1 - r^m}{1 - r}$$

therefore the series converges to

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1 - r}$$

when $|r| < 1$, otherwise it diverges.

Taylor Series I

- Another important approximation in mathematics is approximating functions by their Taylor series expansion.
- Taylor series expanses play an important role in asymptotic statistics.

Taylor Series II

Definition (Taylor Series from Wikipedia)

The Taylor series of a real or complex valued function $f(x)$ that is infinitely differentiable at a real or complex value a is the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

where $n!$ is the factorial of n and $f^{(n)}(a)$ is the n^{th} derivative of $f(x)$ evaluated at the point a.

Taylor Series III

- One of the beauties of the Taylor series is that it can be used to approximate functions at a point.
- To approximate a function we take the first few terms of the Taylor series and consider the rest as being some 'small' remainder.