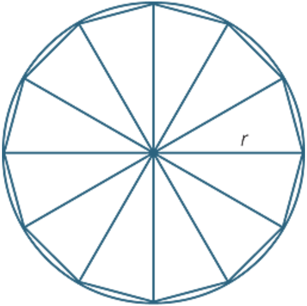
MATH7501: Mathematics for Data Science I

Unit 5: Sequences, Limits and Series

Slides by Timothy Buttsworth (2021)

Motivation: The Need for Sequences

What is the area of a circle of radius 1? By definition (almost), the answer is π, but what is the actual number?



http://www.amsi.org.au/teacher_modules/the_circle.html

We can approximate this number, perhaps using geometry, or other sophisticated methods.

Motivation: The Need for Sequences

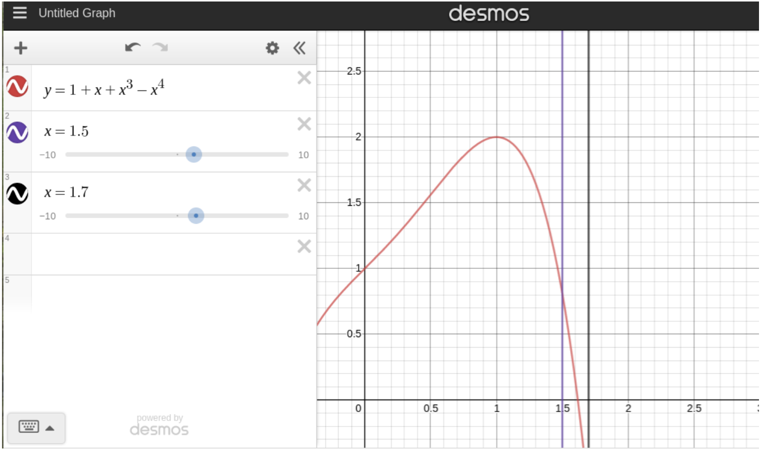
Approximations for π:

- Babylonians, $3, 3.125$;
- Lui Hui 263CE, 3.14;
- Aryabhata 6th century, 3.1416;
- Jamshid al-Kashi 1424, 3.14159265358979324;
- Emma Haruka Iwao 2019, correct to 31.4 trillion digits. She was an employee of Google, and did the approximation over 121 days using Google cloud machines.

As technology improves, we will get arbitrarily good approximations of π, but we will never know what it is exactly!

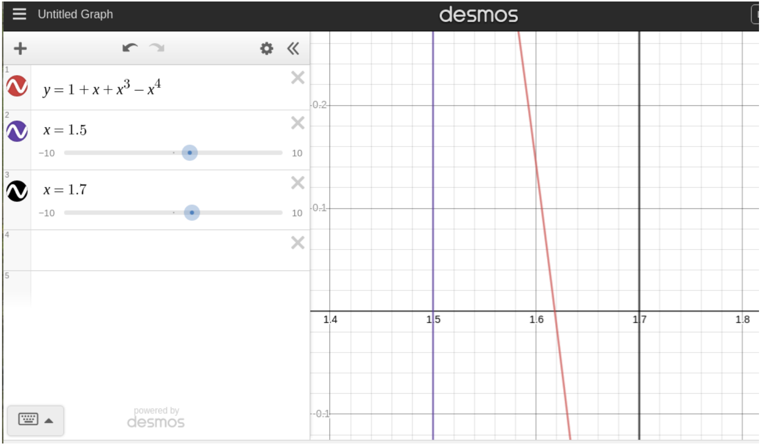
Motivation: The Need for Sequences

Consider the function $f(x) = 1 + x + x^3 - x^4$. Find a positive x so that $f(x) = 0$.



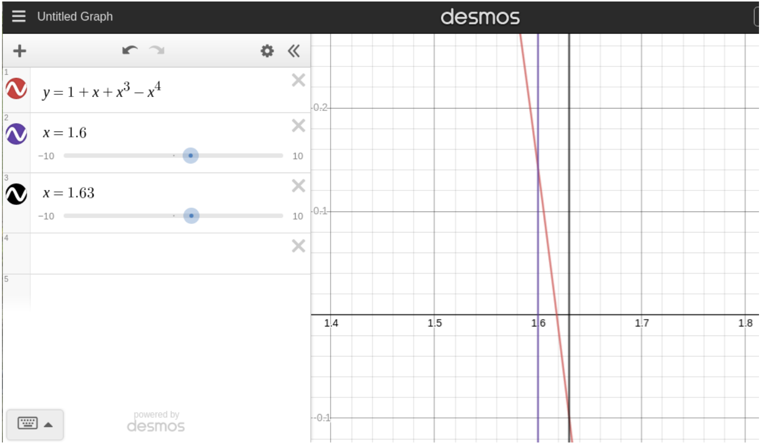
Motivation: The Need for Sequences

Consider the function $f(x) = 1 + x + x^3 - x^4$. Find a positive x so that $f(x) = 0$.



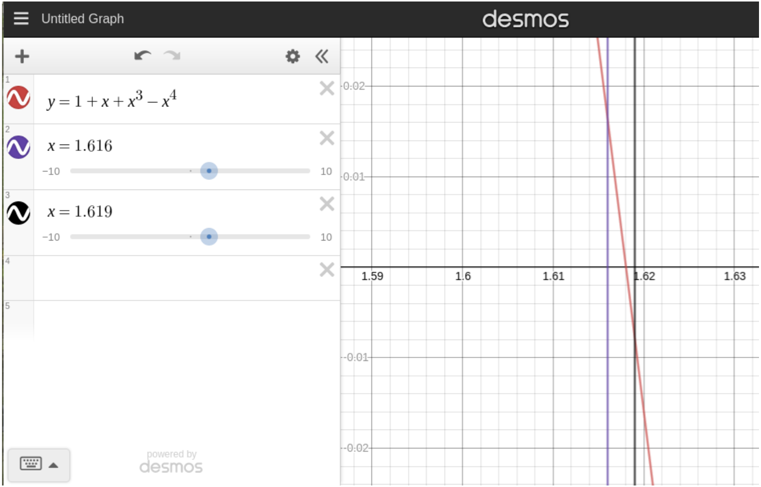
Motivation: The Need for Sequences

Consider the function $f(x) = 1 + x + x^3 - x^4$. Find a positive x so that $f(x) = 0$.



Motivation: The Need for Sequences

Consider the function $f(x) = 1 + x + x^3 - x^4$. Find a positive x so that $f(x) = 0$.



Motivation: The Need for Sequences

Some things to note about these approximations:

- The numbers $1.5, 1.6, 1.616$ are all lower than the true solution, but are getting close.
- The numbers $1.7, 1.63, 1.619$ are all higher than the true solution, but are also getting close.
- We could continue in this manner, with accuracy increasing.
- The true solution is $\frac{1+\sqrt{5}}{2}$, but $\sqrt{5}$ is like π, in the sense that we cannot describe it exactly.

5.1 Sequences of Real Numbers

If we have a large collection of real numbers (possibly infinite), and their order is important, they can be stored in a sequence.

Definition (Sequences)

A *sequence* is an ordered set of real numbers. Suppose the ordered numbers are $a_1, a_2, a_3, \cdots, a_n$ (n could be ∞).Then:

- The sequence is typically denoted with $\{a_k\}_{k=1}^{n}$ (n could be ∞).
- Each individual number in the sequence is referred to as a *term* of the sequence; the a_i number is referred to as the *ith term of the sequence* (so a_1 is the first term of the sequence).

We do not insist that the real numbers are subscripted with 1 at the start. For example, we could have a sequence $a_0, a_1, a_2, \cdots$; this time, a_i is the $i + 1$th term of the sequence.

Example (Digits of π)

Let $\{a_k\}_{k=0}^{\infty}$ be the sequence so that a_k is π correct to the first k decimal places. So the sequence is

$$3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415, \cdots$$

The first term of the sequence is 3, the second term of the sequence is 3.1 and so on. Although there are infinitely-many terms in this sequence, we can only write down the ones for $k \leq 31.4$ trillion.

Example (Ancestors)

Let $\{a_k\}_{k=1}^{\infty}$ be the sequence so that a_k is is the number of ancestors you have at a generation k above you. That is, a_1 is the number of parents you have, a_2 is the number of grand-parents you have, and so on. Then the sequence $\{a_k\}_{k=1}^{\infty}$ looks like

$$2,\ 4,\ 8,\ 16,\ \cdots$$

In this case, we see that

$$a_k = 2^k \text{ for each } k.$$

This is an example of a *general formula for a sequence*. These arise any time we can write an arbitrary term a_k explicitly in terms of k.

5.1 Sequences of Real Numbers

Question

Suppose we have a sequence $\{a_k\}_{k=1}^{\infty}$ described explicitly by the following general formula:

$$a_k = (-1)^{k+1}\frac{k}{k+3}.$$

Compute the first four terms of this sequence.

Solution

- The first term is $a_1 = (-1)^{1+1}\frac{1}{1+3} = \frac{1}{4}$.
- The second term is $a_2 = (-1)^{2+1}\frac{2}{2+3} = -\frac{2}{5}$.
- The third term is $a_3 = (-1)^{3+1}\frac{3}{3+3} = \frac{3}{6} = \frac{1}{2}$.
- The fourth term is $a_4 = (-1)^{4+1}\frac{4}{4+3} = -\frac{4}{7}$.

Question

Find a general formula for a sequence $\{a_k\}_{k=1}^{\infty}$ whose first five terms are $-1, 4, -27, 256, -3125$.

Solution

The alternating minus signs means the sequence can look like

$$a_k = (-1)^k b_k,$$

for some sequence of positive numbers $\{b_k\}_{k=1}^{\infty}$ with $b_1 = 1, b_2 = 4, b_3 = 27, b_4 = 256, b_5 = 3125.$

The b_k sequence can be $b_k = k^k$, so

$$a_k = (-1)^k k^k$$

is a possibility.

Sigma Sum Notation

Suppose we have a sequence

$$a_1,\ a_2,\ a_3, \cdots$$

and we wanted to add up all of the terms between a_m and a_n (inclusive). If $m < n$, we denote this with

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + \cdots + a_{n-1} + a_n.$$

The letter i is called a *dummy variable* because it can be replaced with any other letter, and the expression would still mean the same thing. On the other hand, the choice of n and m affect this expression as they tell us where to start and stop the adding.

5.1 Sequences of Real Numbers

In Mathematica, use the following:

```
In[2]:= Sum[1 / n^2, {n, 1, 100}]

Out[2]= 1 589 508 694 133 037 873 112 297 928 517 553 859 702 383 498 543 709 859 889 432 834 803 818 131 090 ⋱
          369 901 /
        972 186 144 434 381 030 589 657 976 672 623 144 161 975 583 995 746 241 782 720 354 705 517 986 ⋱
          165 248 000
```

Use `1.` to get the decimal representation:

```
In[3]:= Sum[1. / n^2, {n, 1, 100}]

Out[3]= 1.63498
```

Question (Convert Sigma Sum to Expanded Notation)

Write out the summation $\sum_{i=1}^{n}(2i-1)$ in expanded form by writing out the first five terms and the last term.

Solution

$$\sum_{i=1}^{n}(2i-1) = 1+3+5+7+9+\cdots+(2n-1).$$

Question (Convert Expanded to Sigma Sum Notation)

Express the sum

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots+\frac{1}{n\cdot(n+1)}$$

in Sigma Sum notation.

Solution

Terms appear to be of the form $\frac{1}{i(i+1)}$ for i running from 1 to n, so

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots+\frac{1}{n\cdot(n+1)}=\sum_{i=1}^{n}\frac{1}{i(i+1)}.$$

Question (Sum of Squares)

It is well-known that $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. Use this expression to find a simplified expression for $\sum_{i=1}^{n+1} i^2$ in terms of n.

Solution

Since $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$ holds for each n, we can replace n with $n+1$ to find

$$\begin{aligned}\sum_{i=1}^{n+1} i^2 &= \frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}\\ &= \frac{(n+1)(n+2)(2n+3)}{6}.\end{aligned}$$

5.1 Sequences of Real Numbers

Product Notation

For a given sequence $\{a_k\}$, the expression $\prod_{k=m}^{n} a_k$ means

$$a_m \times a_{m+1} \times \cdots \times a_n,$$

i.e., $\prod_{k=m}^{n} a_k$ is the product of the terms in the sequence between a_m and a_n (with $n > m$).

Definition (Factorials)

For a given positive integer n, the number $n!$ (pronounced 'n factorial') is the product of all positive integers up to and including n, i.e.,

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1.$$

By convention, $0! = 1$.

Question

Simplify the following expressions as much as possible:

$$\prod_{j=2}^{5} 2^j, \qquad \frac{10!}{8!2!}, \qquad \frac{(n+2)!}{n!}.$$

Solution

- $\prod_{j=2}^{5} 2^j = 2^2 \times 2^3 \times 2^4 \times 2^5 = 2^{14} = 16384$
- $\frac{10!}{8!2!} = \frac{10\times9\times8\times7\times\cdots\times2\times1}{(8\times7\times\cdots\times2\times1)\times(2\times1)} = \frac{10\times9}{2\times1} = 45$
- $\frac{(n+2)!}{n!} = \frac{(n+2)\times(n+1)\times n\times\cdots\times2\times1}{n\times(n-1)\times\cdots\times2\times1} = (n+2)(n+1)$

Definition (Recursion Relation)

A sequence $\{a_k\}$ so that each term can be written explicitly as a function of the preceding terms is known as a *recursive sequence*, and the explicit formula itself is known as a *recursion relation*.

Example

Consider the recursive sequence $\{a_k\}_{k=1}^{\infty}$ so that $a_1 = 2$, and for each $k \geq 1$,

$$a_{k+1} = \frac{1}{3 - a_k} \text{ (recursive formula).}$$

The first four terms of this sequence are $2, 1, \frac{1}{2}, \frac{2}{5}$.

Example (The Fibonacci Sequence)

The Fibonacci sequence is a sequence of integers $\{F_k\}_{k=1}^{\infty}$ so that

$$\begin{aligned} F_1 &= 1 \\ F_2 &= 1 \\ F_k &= F_{k-1} + F_{k-2}, \ \textit{for } k \geq 3. \end{aligned}$$

The first 8 terms are $1, 1, 2, 3, 5, 8, 13, 21$. It turns out there is a general formula for the Fibonacci sequence:

$$F_k = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^k - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^k .$$

This can be proven using strong mathematical induction.

Some Remarks About Sequences

So far we have seen three ways of describing sequences:

1. Writing out the first few terms in the sequence, and 'hoping' that the general pattern is obvious. For example, consider the sequence whose first few terms are $3, 5, 7, \cdots$. What is the next term?
2. Giving a general formula which explains how each term is calculated. For example, consider the sequence $\{a_n\}_{n=0}^{\infty}$ so that $a_n = \frac{(-1)^n}{n+2}$.
3. Writing out the first few terms, and then giving a formula which explains how any terms can be constructed from the preceding terms. For example, a sequence $\{c_i\}_{i=0}^{\infty}$ with $c_0 = 1$, $c_1 = 2$ and $c_s = c_{s-1} + 3c_{s-2} + 1$ for all $s \geq 2$.

It is sometimes possible to move between descriptions for a given sequence.

Question

Let $\{c_k\}_{k=0}^{\infty}$ be a sequence so that

$$c_k = (k-1)c_{k-1} + kc_{k-2} + k, \text{ for } k \geq 2$$
$$c_0 = 1, \qquad c_1 = 2.$$

Calculate c_2, c_3, c_4.

Solution

- $c_2 = (2-1)c_1 + 2c_0 + 2 = 2 + 2 + 2 = 6.$
- $c_3 = (3-1)c_2 + 3c_1 + 3 = 12 + 6 + 3 = 21.$
- $c_4 = (4-1)c_3 + 4c_2 + 4 = 63 + 24 + 4 = 91.$

Question

Let $\{b_i\}_{i=0}^{\infty}$ be a sequence so that

$$b_i = 5b_{i-1} - 6b_{i-2}, \text{ for } i \geq 2,$$

with b_0, b_1 unspecified. Find expressions for b_{i+1}, b_{i+2}.

Solution

The given formula holds for all $i \geq 2$, so

$$b_{i+1} = 5b_{(i+1)-1} - 6b_{(i+1)-2} = 5b_i - 6b_{i-1}.$$

Similarly,

$$b_{i+2} = 5b_{(i+2)-1} - 6b_{(i+2)-2} = 5b_{i+1} - 6b_i.$$

Question

Consider the sequence $\{a_n\}_{n=0}^{\infty}$ with $a_n = 5 \cdot 2^n$. Show that this can be defined recursively with

$$\begin{aligned} a_n = 2a_{n-1}, \ \text{for } n \geq 1, \\ a_0 = 5. \end{aligned}$$

Solution

Clearly $a_0 = 5$ and $a_1 = 10$. Also by definition of the sequence,

$$a_n = 5 \cdot 2^n = 2 \cdot (5 \cdot 2^{n-1}) = 2a_{n-1}.$$

Question

The Catalan numbers can be defined as $C_n = \frac{1}{n+1}\binom{2n}{n}$ for all integers $n \geq 1$. Find C_i for $i = 1, 2, 3$, and then show that $C_k = \frac{4k-2}{k+1} C_{k-1}$ for all $k \geq 2$.

Solution

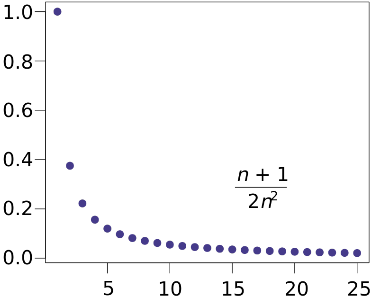
$C_1 = \frac{2!}{2 \cdot 1! \cdot 1!} = 1$, $C_2 = \frac{4!}{2! \cdot 2! \cdot 3} = 2$, $C_3 = \frac{6!}{3! \cdot 3! \cdot 4} = 5$. In general

$$\begin{aligned} C_n &= \frac{2n!}{n!n!(n+1)} \\ &= \frac{2n(2n-1)}{(n+1)n} \cdot \frac{(2(n-1))!}{n(n-1)!(n-1)!} \\ &= \frac{4n-2}{n+1} C_{n-1}. \end{aligned}$$

5.2 Limits of Sequences

Definition (Limits)

Consider a sequence $\{a_n\}_{n=0}^{\infty}$. We say that the sequence *converges to the number l* if the terms a_n can be made arbitrarily close to l by having n arbitrarily large. More formally, the sequence converges to l if for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ so that $|a_n - l| < \epsilon$ whenever $n \geq N$.



https://en.wikipedia.org/wiki/Limit_of_a_sequence

Conventions (Limits)

If the sequence $\{a_n\}_{n=1}^{\infty}$ converges to the number l, then l is said to be the *limit* of the sequence, written $l = \lim_{n\to\infty} a_n$. In this case, the sequence is said to be *convergent*. A sequence which does not have a limit is said to be *divergent*.

5.2 Limits of Sequences

Question

Determine convergence/divergence of the following sequences. Find the limits (if they exist).

- $a_n = \frac{1}{n}$;
- b_n defined recursively with $b_0 = -1$, $b_n = \frac{b_{n-1}}{2}$ for $n \geq 1$.

Solution

- The terms $\frac{1}{n}$ become arbitrarily small by having n large, so $\lim_{n\to\infty} a_n = 0$. The sequence is therefore convergent.
- Converting to direct form, we find $b_n = -\frac{1}{2^n}$. The terms $\frac{-1}{2^n}$ become arbitrarily small by having n large, so $\lim_{n\to\infty} b_n = 0$. The sequence is convergent.

5.2 Limits of Sequences

Question

Determine convergence/divergence of the following sequences. Find the limits (if they exist).

- $c_n = (-1)^n$;
- $d_n = r^n$ for some fixed real number r.

Solution

- Since $(-1)^n$ is 1 for even n and -1 for odd n, the sequence $\{c_n\}$ oscillates repeatedly between -1 and 1. While the terms get close to both -1 and 1, the sequence does not **stay** close to either one of them, so the sequence is divergent.
- We have sub cases: for $\{d_n\}$:
  - If $|r| < 1$ the terms shrink and $\lim_{n\to\infty} d_n = 0$.
  - If $r = 1$, $d_n = 1$ for all n, so $\lim_{n\to\infty} d_n = 1$.
  - If $r = -1$, then $d_n = (-1)^n = c_n$, so this sequence is divergent.
  - If $|r| > 1$, then the size of the terms grows, becoming arbitrarily large, so the sequence is divergent.

5.2 Limits of Sequences

Theorem (Properties of Sequences)

Suppose that $\{a_n\}$ and $\{b_n\}$ are convergent sequences with $\lim_{n\to\infty} a_n = l$ and $\lim_{n\to\infty} b_n = m$. Then for any constant $c \in \mathbb{R}$, the following are true:

- $\lim_{n\to\infty}(a_n \pm b_n) = l \pm m$;
- $\lim_{n\to\infty} ca_n = cl$;
- $\lim_{n\to\infty}(a_n b_n) = lm$;
- $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{l}{m}$, provided $m \neq 0$.

Remark (Dividing by Sequences Converging to 0)

If $m = 0$, then really anything could happen in the last point. For example, we could take the convergent sequences $b_n = \frac{1}{n^2}$ and $a_n = \frac{1}{n}$. Then $\frac{a_n}{b_n} = n$, so $\{\frac{a_n}{b_n}\}$ is divergent.

On the other hand, if $b_n = \frac{1}{n^2}$ and $a_n = \frac{1}{n^2}$, then $\frac{a_n}{b_n} = 1$ for all n, so the sequence converges to 1.

5.2 Limits of Sequences

Theorem (Squeeze Theorem)

Suppose that $\{a_n\}, \{c_n\}$ are sequences, both convergent to l. If $\{b_n\}$ is a sequence so that there is an $n_0 \in \mathbb{N}$ so that $a_n \leq b_n \leq c_n$ whenever $n \geq n_0$, then $\lim_{n\to\infty} b_n = l$.

Example

Let $a_n = \frac{\sin(n)}{n}$. Then $-\frac{1}{n} \leq a_n \leq \frac{1}{n}$. Both outer sequences converge to 0, so $\lim_{n\to\infty} a_n = 0$.

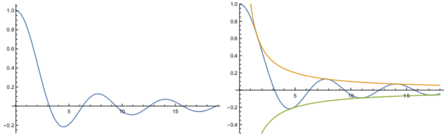


Figure 37: A plot of $f(x) = \frac{\sin(x)}{x}$ and $\pm\frac{1}{x}$ together with $f(x)$.

5.2 Limits of Sequences

Examples (Well-Known Limits)

- For a constant c, $\lim_{n\to\infty} c^n = \begin{cases} 0 \text{ if } |c| < 1, \\ 1 \text{ if } c = 1. \end{cases}$ The sequence is divergent for other c.
- For constant $c > 0$, $\lim_{n\to\infty} c^{\frac{1}{n}} = 1$.
- $\lim_{n\to\infty} \frac{1}{n^r} = 0$ for any $r > 0$.
- $\lim_{n\to\infty} n^{\frac{1}{n}} = 1$.
- $\lim_{n\to\infty} \frac{1}{n!} = 0$.
- For any constant c, $\lim_{n\to\infty} \frac{c^n}{n!} = 0$.
- $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e$.
- For any real a, $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$.

5.3 Series

Definition (Infinite Sum)

Let $\{a_i\}_{i=1}^{\infty}$ be a sequence. The *series* of $\{a_i\}$ is the new sequence $\{s_n\}_{n=1}^{\infty}$ so that $s_n = \sum_{i=1}^{n} a_i$, i.e., the first few terms of $\{s_n\}$ are

$$a_1,\ a_1 + a_2,\ a_1 + a_2 + a_3,\ a_1 + a_2 + a_3 + a_4, \cdots .$$

The term s_n is called the nth *partial sum*.

The limit of this sequence is called the *infinite sum of* $\{a_n\}$, and is denoted $\sum_{i=1}^{\infty} a_i$. In the case the limit exists, we say that the series *converges*, and if the limit does not exist, we say the series *diverges*.

Note that the original sequence does not necessarily have to start at $i = 1$.

Mathematica can sometimes compute infinite sums:

```
In[1]:= Sum[1/n^2, {n, 1, Infinity}]
```

Out[1]= $\frac{\pi^2}{6}$

5.3 Series

Applications in Infinite Sums

- The number π can be expressed as the infinite sum $3 + 0.1 + 0.04 + 0.001 + 0.0005 + \cdots$. More generally, solutions to equations can often be approximated by adding smaller and smaller numbers; the infinite sum should be the true solution.
- Exponential and trigonometric functions can be expressed as infinite sums of polynomials:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

Question

Does $\sum_{n=0}^{\infty}(-1)^n$ converge or diverge?

Solution

$$\begin{aligned} s_0 &= 1 \\ s_1 &= 1 + (-1) = 0 \\ s_2 &= 1 + (-1) + 1 = 1 \\ &\vdots \end{aligned}$$

and this pattern repeats. In general, $s_{2m} = 1$ and $s_{2m+1} = 0$, so the limit of $\{s_n\}$ does not exist, and the infinite sum diverges.

Question (An example of a telescoping series)

Does $\sum_{n=0}^{\infty} \frac{1}{(n+1)(n+2)}$ converge or diverge?

Solution

First, note that $\{\frac{1}{(n+1)(n+2)}\}_{n=1}^{\infty}$ is the same as $\{\frac{1}{n(n+1)}\}_{n=0}^{\infty}$. So it suffices to examine convergence of $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$.

We compute the partial sums:

$$s_n = \sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \frac{1}{i} - \frac{1}{i+1}$$
$$= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right).$$

We notice some cancellation, so $s_n = 1 - \frac{1}{n+1}$, and $\lim_{n\to\infty} s_n = 1$; the series converges.

Theorem (p-test for series)

For any $p \in \mathbb{R}$, the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is convergent whenever $p > 1$, and is divergent is $p \leq 1$.

Theorem (n-term test)

If the series $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

CAUTION: this Theorem can only ever be used to conclude that a series diverges. It can NEVER be used to conclude that a series converges.

Example (Harmonic Series)

The harmonic series is $\sum_{n=1}^{\infty} \frac{1}{n}$. By the p-test, this series is divergent, even though $\lim_{n\to\infty} \frac{1}{n} = 0$ (an example which shows why you can never use the nth term test to conclude that a series converges).

To see why this series diverges, we can write the sum as

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots \\ &\geq 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots .\end{aligned}$$

So no matter how far you go, there will always be another $\frac{1}{2}$, so the series is divergent.

5.3 Series

Theorem (Geometric Series)

If a is a real number and $|r| < 1$, then $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$. The series is divergent if $|r| \geq 1$.

Example (Negative Interest Rate)

Suppose that at the start of each year, a person puts 100 into an account which charges 5% interest p.a. Then at the start of the $n+1$th year, the amount in their account is

$$100 + 0.95(100 + 0.95(\cdots)) = \sum_{i=0}^{n} 100 \cdot 0.95^i.$$

As $n \to \infty$, this number increases to $\frac{100}{1-0.95} = \frac{10000}{5} = 2000$.

5.4 Application: Set Cover Problem

Problem (The Set Cover Problem)

Let $U = \{1, 2, \cdots, n\}$, and consider a collection of subsets $S \subseteq 2^U$ so that U is the union of all sets in S. What is the smallest number of sets in S needed so that the union of these sets is still equal to U?

Example

Let $U = \{1, 2, 3, 4, 5\}$, and consider a collection of subsets

$$S = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}\}.$$

Then U can be covered with only to elements of S, namely $\{1, 2, 3\}$ and $\{4, 5\}$.

5.4 Application: Set Cover Problem

How do we find these sets generally?

The Greedy Algorithm

At each stage, include the subset in S that contains the largest number of elements of U that have not already been covered.

Example

Let $U = \{1, 2, 3, 4, 5\}$, and consider a collection of subsets

$$S = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}\}.$$

Then the first set to be chosen is $\{1, 2, 3\}$, and the second set is $\{4, 5\}$, so we are done!

This algorithm does not always give you the correct answer! For example, if $U = \{1, 2, 3, 4, 5, 6\}$, and

$$S = \{\{2, 3, 4, 5\}, \{1, 2, 3\}, \{4, 5, 6\}\}$$

then the greedy algorithm will select all three sets. However, only $\{1, 2, 3\}$ and $\{4, 5, 6\}$ are needed.

In general, the greedy algorithm will produce an answer that could be as large as $H(n)$ times the true answer, where $H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is the nth harmonic number. Recall that the corresponding infinite sum diverges, so there is no limit to just how poor the answer could be!