Dedekind's forgotten axiom and why we should teach it (and why we shouldn't teach mathematical induction in our calculus classes)

by Jim Propp (UMass Lowell)

March 14, 2010

Completeness

Three common axioms of completeness, and one not-so-common:

- convergence of all Cauchy sequences
- nested interval property
- what people call "Dedekind completeness"
- what people *should* call "Dedekind completeness"

What people usually call Dedekind completeness:

Every bounded non-empty set S of real numbers has a least upper bound.

It's a good completeness axiom, but it's not in Dedekind!

Question: Who deserves the credit for this axiom?

What Dedekind did in his 1872 pamphlet “Continuity and irrational numbers” (section V, subsection IV, in Beman’s translation, taken from the Dover paperback *Essays on the Theory of Numbers*):

(a) stated completeness of line (more on this below)

(b) defined Dedekind cuts to build a model of **R** from **Q**

(c) derived completeness of **R** (more on this below)

(a) "If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions. ... The assumption of this property of the line is nothing else than an axiom by which we attribute to the line its continuity."

Note that Dedekind used the word *continuity* (*Stetigkeit*) where we nowadays would say *completeness*.

(b) [Details well-known and hence omitted.]

(c) "The following theorem is true: If the system **R** of all real numbers breaks up into two classes A_1, A_2 such that every number α_1 of the class A_1 is less than every number α_2 of the class A_2 then there exists one and only one number α by which this separation is produced."

(Earlier he explains how, given a number α, one can produce sets A_1, A_2 satisfying this condition, and indeed can do so in two ways for each α, according to whether α belongs to A_1 or A_2.)

Note that for Dedekind, *set* means *nonempty set*.

I'll call (c) the *cut property of* $\mathbf{R}$:

The cut property: If A, B are non-empty subsets of $\mathbf{R}$ such that $A \cup B = \mathbf{R}$, $A \cap B = \emptyset$, and $a < b$ for all $a \in A$ and $b \in B$, then there exists c in $\mathbf{R}$ such that $a \leq c$ for all $a \in A$ and $c \leq b$ for all $b \in B$.

More compactly:

Every *cut* of $\mathbf{R}$ is associated with an *element* of $\mathbf{R}$.

The cut property is a second-order property, like the least upper bound property (though the collection of sets we're quantifying over is easier to visualize, and in fact turns out to have smaller cardinality).

Main theme: We could bypass the introduction of Dedekind cuts and just take the cut property of $\mathbf{R}$ as an axiom.

The cut property implies the least upper bound property, and vice versa:

$\Rightarrow$: Every non-empty set S that's bounded above cuts $\mathbf{R}$ into two pieces: the set B of numbers that are upper bounds of S and the set A of numbers that aren't. The cut-point c given by the cut property can be shown to be the least upper bound of S.

$\Leftarrow$: Take c to be the least upper bound of A.

The cut property is simple and compelling.

It is also satisfyingly symmetrical (unlike the least upper bound property, whose twin, the greatest lower bound property, is logically equivalent but semantically different).

Why don't we call the cut property an axiom of real analysis?

The cut property is easier for students to grasp than the least upper bound property.
(It's easier for students to visualize a cut of $\mathbf{R}$ than an arbitrary non-empty bounded-above subset of $\mathbf{R}$.)

The cut property is teachable Socratically.
(Ask your students to come up with a set $\emptyset \subset B \subset \mathbf{R}$ such that every non-element of B is less than every element of B. Then ask them to find another such set. And then another. Then challenge them to find *all* such sets. When they've come up with all the sets $B = [c, \infty)$ and $B = (c, \infty)$ and can't think of any more, they'll be ready to conjecture the cut property on their own.)

Why don't we teach it?

“Forgotten”?

The geometrical version was never forgotten by workers in the foundations of geometry: see e.g. Marvin Jay Greenberg’s “Old and New Results in the Foundations of Elementary Plane Euclidean and Non-Euclidean Geometry”, American Mathematical Monthly **117**, Number 3, March 2010, pages 198–219.

Logicians didn't forget: Tarski's continuity axiom (If A, B are subsets of $\mathbf{R}$ such that $a < b$ for all $a \in A$ and $b \in B$, then there exists $c \in \mathbf{R}$ such that that $a \leq c$ for all $a \in A$ and $c \leq b$ for all $b \in B$) is a variant of Dedekind's (logically equivalent but semantically stronger since its hypotheses are weaker).

Historians didn't forget: see Steven Krantz's on-line book `www.math.wustl.edu/~sk/books/newhist.pdf`, *An Episodic History of Mathematics*, page 353.

Some analysts remember: Dedekind's cut property appears as Theorem 1.32 in Walter Rudin's *Principles of Mathematical Analysis*.

But I've found only one intro textbook on real analysis that takes this property of $\mathbf{R}$ as a starting point: *Real and Complex Analysis* by C. Apelian and S. Surace.

Why doesn't everyone base real analysis on the cut property?

I have a couple of ideas about this:

How ideas are originally labelled and packaged makes a difference:

Dedekind treated the property as a *theorem*, so it takes intellectual effort for readers of Dedekind to recognized it as a good *axiom*

Dedekind didn't give the property a name

Dedekind didn't state the property in a succinct self-contained way (and it's somewhat resistant to being stated in such a fashion)

How ideas lend themselves to use makes a difference:

There's nothing that the cut axiom can do that the least upper bound axiom can't do better

There isn't time in the semester to teach a principle you'll only use once

From completeness to induction

The usual approach to proving formulas like

$$\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$$

is via the axiom of induction:

If a proposition is true for 1, and if whenever the proposition is true for n it's true for $n+1$, then the proposition is true for all positive integers.

When we teach our students to prove such formulas in this way, we're missing a huge opportunity to convey a perspective that unifies the discrete and continuous realms.

The first hint of this comes from the observation that the axiom of induction for $\mathbf{N}$ can be derived as a consequence of the completeness of $\mathbf{R}$ by way of

- the ordered field axioms for $\mathbf{R}$, and
- the ordered ring axioms for $\mathbf{Z}$, and
- one extra axiom: the axiom that 1 is the smallest positive integer.

This implication may seem less bizarre if one considers some parallels between various statements equivalent to completeness and various statements equivalent to induction.

Greatest lower bound property	Least element principle
Topological connectedness of **R**	Graph-theoretic connectedness of **Z**
Cut axiom for **R**	Cut axiom for **Z**

Greatest lower bound property: A non-empty subset of $\mathbf{R}$ that's bounded below has a greatest lower bound.

Least element principle: A non-empty subset of $\mathbf{Z}$ that's bounded below has a least element.

Topological connectedness of $\mathbf{R}$: $\mathbf{R}$ cannot be expressed as the disjoint union of two topologically closed non-empty subsets.

Graph-theoretic connectedness of $\mathbf{Z}$: $\mathbf{Z}$ cannot be expressed as the disjoint union of two graph-theoretically closed non-empty subsets.

(A subset S of the vertex set V of a graph is closed iff there exist no edge of the graph joining a vertex of S to a vertex of $V \setminus S$.)

Cut axiom for **R**: If A, B are disjoint non-empty subsets of **R** with every element of A less than every element of B, then A has a greatest element or B has a least element.

(This is equivalent to what I called the cut axiom before, as long as you're not an intuitionist.)

Cut axiom for **Z**: If A, B are disjoint non-empty subsets of **Z** with every element of A less than every element of B, then A has a greatest element and B has a least element.

(Equivalently: ... then there exist $a \in A$, $b \in B$ with $b = a + 1$.)

Question: Can one derive the cut axiom for **R** from the cut axiom for **Z** via non-standard analysis?

But the best pair (of a completeness principle for **R** and an induction principle for **N** with a strong thematic link between them) — best conceptually *and* best pedagogically — is

Continuous no-change-implies-constancy principle	**Discrete** no-change-implies-constancy principle

Instead of teaching the axiom of induction, I propose we teach calculus students the following axiom:

The discrete no-change-implies-constancy principle: If $f : \mathbf{N} \to \mathbf{R}$ has the property that $f(n+1) - f(n) = 0$ for all n in $\mathbf{N}$, then f is constant on $\mathbf{N}$.

This is indeed how computers generate proofs of identities nowadays, thanks to the work of Herb Wilf and Doron Zeilberger.

Their method involves a number of tricks, but the basic trick is: to prove that $a(n) = b(n)$ for all $n \geq 1$, first show that $a(n) - b(n)$ is constant (by showing that it doesn't change if you replace n by $n+1$), and then evaluate the constant.

Easy exercise: Apply this to $a(n) = 1^1 + 2^2 + \cdots + n^2$ and $b(n) = n(n+1)(2n+1)/6$.

The discrete no-change-implies-constancy principle implies the axiom of induction, and vice versa.

$\Leftarrow$: [Omitted.]

$\Rightarrow$: Given propositions $P(1)$, $P(2)$, …such that $P(1)$ is true and such that (for all n) $P(n)$ implies $P(n+1)$, let

$$f(n) = \begin{cases} 1 & \text{if } P(1) \text{ through } P(n) \text{ are true,} \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to show that $f(n) = f(n+1)$ for all n, so f is constant. And since $f(1) = 1$, that constant is 1, so $f(n) = 1$ for all n, so $P(n)$ is true for all n.

The continuous no-change-implies-constancy principle: If $f : \mathbf{R} \to \mathbf{R}$ is differentiable and satisfies $f'(x) = 0$ for all $x \in \mathbf{R}$, then f is constant on $\mathbf{R}$.

The continuous no-change-implies-constancy principle implies the axiom of completeness, and vice versa.

$\Leftarrow$: Completeness implies the Extreme Value Theorem, which implies the Mean Value Theorem, which implies the continuous no-change-implies-constancy principle.

$\Rightarrow$: We prove the contrapositive. Suppose completeness were false. There would exist a cut of $\mathbf{R}$ into two non-empty sets A and B, with every element of A less than every element of B, where A has no greatest element and B has no least element.
Then the indicator function of A would be a non-constant function whose derivative is everywhere zero.
So the continuous no-change-implies-constancy principle would be false.

Note that I am **not** suggesting that we should take the continuous no-change-implies-constancy principle as an axiom.

For one thing, we can't even state it until we've defined the derivative.

But most calculus texts already mention the principle.

We can emphasize it more.

Then, in the second semester, when we need formulas like $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$, we can bring in the discrete no-change-implies-constancy principle, and it'll make sense because it's analogous to something the students have already seen.

Slides for this talk are at
`http://jamespropp.org/hpm10-slides.pdf`

An early version of this talk (after I'd reinvented Dedekind's cut property but before I learned it was Dedekind's) is at
`http://jamespropp.org/cut.pdf`

I'm thinking of writing this up for publication; comments are welcome!