University of Wyoming
Math 2800—Fall 2012
Department of Mathematics
Math Majors Seminar

# Sets and Cardinality (Handout September 25, 2012)

Imagine an illiterate (and innumerate) shepherd checking to see that all his sheep are present. He doesn't know how many sheep he has; he doesn't even know names for numbers. Instead he keeps a set of pebbles in a pouch. He knows that there is one pebble for each sheep. To check that all his sheep are present, he passes through the herd, slipping one pebble from his palm back into the pouch every time he passes a sheep. This story teaches us an important lesson about counting: the principle of finding a one-to-one correspondence between two sets (in this case a herd of sheep and a collection of pebbles) is more basic than using names of numbers (0, 1, 2, 3, ...) to establish the size of set. Having names for numbers might, of course, be useful later if the shepherd wants to tell someone else how many sheep he has.

This insight motivates our approach to comparing sizes of sets in all cases, including the case of infinite sets. Recall that a *set* is a collection of objects. To properly define the concept of a set requires a little more care than this; but let's not worry about such technicalities right now. To compare the size of two sets $S$ and $T$, we try to establish a one-to-one correspondence between their elements. The first cardinal rule is:

(C1) If there is a one-to-one correspondence between elements of $S$ and elements of $T$ (i.e. if there exists a bijection $S \to T$), then the sets $S$ and $T$ *have the same cardinality* (i.e. the same size), denoted by $|S| = |T|$.

We are not yet giving a name to the cardinality of a set; we only say what it means for two sets to have the same size. According to this definition, the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ and the set of counting numbers $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$ have the same size; this is because there is a one-to-one correspondence between their elements given by

$$\begin{array}{rccccc} \mathbb{N} = \{ & 1, & 2, & 3, & 4, & \ldots \} \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{N}_0 = \{ & 0, & 1, & 2, & 3, & \ldots \} \end{array}$$

This might seem disturbing at first, since $\mathbb{N}$ is a proper subset of $\mathbb{N}_0$: every element of $\mathbb{N}$ is an element of $\mathbb{N}_0$, but not conversely. This feature of infinite sets is universal, and may be taken as the definition of an infinite set: A set $S$ is infinite if and only if there is a one-to-one correspondence between $S$ and a proper subset of $S$.

Be careful: there is an obvious correspondence between $\mathbb{N}$ and a subset of $\mathbb{N}_0$, which fails to be a perfect matching:

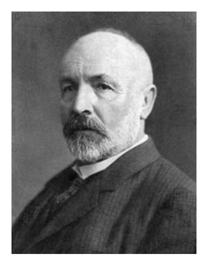
$$\begin{array}{rcccccccccc} \mathbb{N} = \{ & 1, & & 2, & & 3, & & 4, & & \dots \} \\ & & \nwarrow\!\!\!\!\searrow & & \nwarrow\!\!\!\!\searrow & & \nwarrow\!\!\!\!\searrow & & \nwarrow\!\!\!\!\searrow & \\ \mathbb{N}_0 = \{ & 0, & & 1, & & 2, & & 3, & & \dots \} \end{array}$$

This might suggest that $\mathbb{N}$ is smaller than $\mathbb{N}_0$. (*Not!*) Also there is a correspondence between a subset of $\mathbb{N}$ and all of $\mathbb{N}_0$ which is not a perfect matching:

$$\begin{array}{rcccccccccc} \mathbb{N} = \{ & 1, & & 2, & & 3, & & 4, & & \dots \} \\ & & \swarrow\!\!\!\!\nearrow & & \swarrow\!\!\!\!\nearrow & & \swarrow\!\!\!\!\nearrow & & \swarrow\!\!\!\!\nearrow & \\ \mathbb{N}_0 = \{ & 0, & & 1, & & 2, & & 3, & & \dots \} \end{array}$$

This might suggest that on the contrary, $\mathbb{N}_0$ is smaller than $\mathbb{N}$! Neither of these conclusions is correct. Carefully observe that (C1) does not require that *every* correspondence between $S$ and $T$ is a perfect matching. The fact that there is an imperfect matching, doesn't change the fact that there is also a perfect matching. But of course the confusion does not arise for finite sets; and so it is not surprising that this confusion led many leading mathematicians of the early 20th century to conclude that talk about infinite sets was balderdash.

It was Georg Cantor (1845–1918) who first dispelled this pessimistic view by carefully defining cardinality of sets. Sadly, he was ostracized by the mathematical mainstream of his day, and he lost his sanity. In hindsight, his mathematical contributions are considered among the greatest of the 20th century. His rules for comparing the size of sets are (C1) above (to test for equal-size sets) and (C2) below (for sets of possibly differing size).



Georg Cantor
(1845–1918)

The set $\mathbb{Z}$ of all integers, and the set $\mathbb{Q}$ of all rational numbers, both have the same size as $\mathbb{N}$. To see that $|\mathbb{Z}| = |\mathbb{N}|$, observe the one-to-one correspondence:

$$\begin{array}{rcccccccc} \mathbb{N} = \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & \dots \} \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{Z} = \{ & 0, & 1, & -1, & 2, & -2, & 3, & -3, & \dots \} \end{array}$$

The set $\mathbb{Q}$ of all rational numbers takes a little more thought. First list all integers, then all reduced fractions with denominator 2, then those with denominator 3, etc. row-by-row, thus:

$$\begin{array}{cccccccccc}
0, & 1, & -1, & 2, & -2, & 3, & -3, & 4, & -4, & \dots \\
\frac{1}{2}, & -\frac{1}{2}, & \frac{3}{2}, & -\frac{3}{2}, & \frac{5}{2}, & -\frac{5}{2}, & \frac{7}{2}, & -\frac{7}{2}, & \frac{9}{2}, & \dots \\
\frac{1}{3}, & -\frac{1}{3}, & \frac{2}{3}, & -\frac{2}{3}, & \frac{4}{3}, & -\frac{4}{3}, & \frac{5}{3}, & -\frac{5}{3}, & \frac{7}{3}, & \dots \\
\frac{1}{4}, & -\frac{1}{4}, & \frac{3}{4}, & -\frac{3}{4}, & \frac{5}{4}, & -\frac{5}{4}, & \frac{7}{4}, & -\frac{7}{4}, & \frac{9}{4}, & \dots \\
\frac{1}{5}, & -\frac{1}{5}, & \frac{2}{5}, & -\frac{2}{5}, & \frac{3}{5}, & -\frac{3}{5}, & \frac{4}{5}, & -\frac{4}{5}, & \frac{6}{5}, & \dots \\
\frac{1}{6}, & -\frac{1}{6}, & \frac{5}{6}, & -\frac{5}{6}, & \frac{7}{6}, & -\frac{7}{6}, & \frac{11}{6}, & -\frac{11}{6}, & \frac{13}{6}, & \dots \\
\frac{1}{7}, & -\frac{1}{7}, & \frac{2}{7}, & -\frac{2}{7}, & \frac{3}{7}, & -\frac{3}{7}, & \frac{4}{7}, & -\frac{4}{7}, & \frac{5}{7}, & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

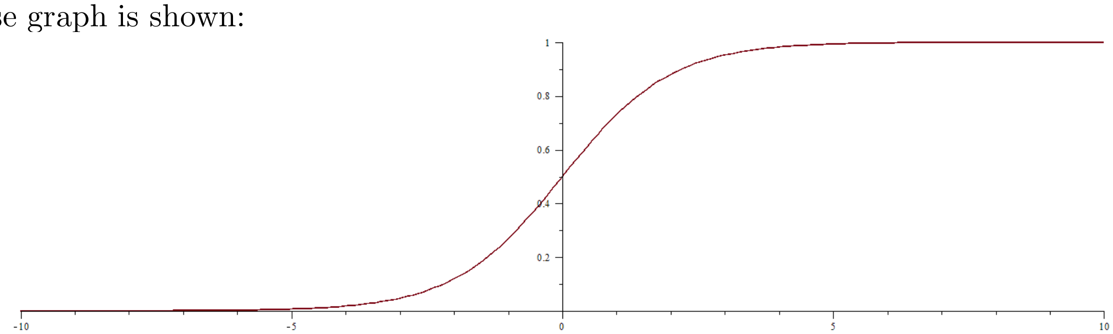
We list all rationals using the snake-like path shown, to obtain a one-to-one correspondence

$$\begin{array}{lcccccccccc}
\mathbb{N} = \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & \dots \} \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\
\mathbb{Q} = \{ & 0, & 1, & \frac{1}{2}, & \frac{1}{3}, & -\frac{1}{2}, & -1, & 2, & \frac{3}{2}, & -\frac{1}{3}, & \dots \}
\end{array}$$

so that all four sets $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{Q}$ have the same size.

A *one-to-one correspondence* or *perfect matching* between two sets $S$ and $T$ is really just a function from $S$ to $T$ that is *bijective*, i.e. both one-to-one and onto. To say that $f : \mathbb{N} \to \mathbb{N}_0$ is *one-to-one* (i.e. *injective*) means that $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$. To say that $f$ is *onto* (i.e. *surjective*) means that for every $y \in \mathbb{N}_0$, there exists $x \in \mathbb{N}$ such that $f(x) = y$. Our bijection $f : \mathbb{N} \to \mathbb{N}_0$ is defined by the formula $f(x) = x - 1$. This function is both one-to-one and onto. Our bijection $g : \mathbb{N} \to \mathbb{Q}$ is not given by any easy formula, but it is both one-to-one and onto; here $g(1) = 0$, $g(2) = 1$, $g(3) = \frac{1}{2}$, ....

Sometimes a one-to-one correspondence is expressible using a graph. Consider, for example, the set $\mathbb{R}$ consisting of all real numbers, and the open interval $(0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$. We have a bijection $f : \mathbb{R} \to (0, 1)$ given explicitly by $f(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$ whose graph is shown:




Since $f$ is an increasing continuous function whose graph has two horizontal asymptotes $y = 0$ and $y = 1$, we may conclude that $f : \mathbb{R} \to (0, 1)$ is in fact bijective. Thus $|\mathbb{R}| = |(0, 1)|$. Observe that the one-to-one correspondence between $x$-values and $y$-values is given by the graph.

It may come as a shock to find that not all infinite sets have the same size; we will show that there is no one-to-one correspondence between $\mathbb{R}$ and $\mathbb{N}$. Since $\mathbb{N} \subset \mathbb{R}$, it is reasonable to expect that $|\mathbb{N}| \leqslant |\mathbb{R}|$. (Recall that having a proper subset does not force $|\mathbb{N}|$ to be smaller than $|\mathbb{R}|$. We have already seen lots of examples to show the fallacy in this reasoning.) We have already invoked the second cardinal rule:

(C2) If $S \subseteq T$ then $|S| \leqslant |T|$. More generally, if $S$ is in one-to-one correspondence with a subset of $T$, then $|S| \leqslant |T|$.

Some remarks about notation: For two sets $S$ and $T$, the statement $S \subseteq T$ (read as '$S$ is a subset of $T$') means that every element of $S$ is an element of $T$ (although elements of $T$ are not assumed to belong to $S$). I will write $S \subset T$ to say that $S$ is a *proper* subset of $T$ (i.e. every element of $S$ is an element of $T$, and at least one element of $T$ is not in $S$). Be careful: some books uses the symbol '$\subset$' to mean '$\subseteq$'. (This confusion has an unfortunate history. Let us hope we never see the day when '$<$' is confused with '$\leqslant$'.)

The following result is due to Cantor.

**Theorem.** $|\mathbb{N}| < |\mathbb{R}|$.

*Proof.* Since $\mathbb{N} \subset \mathbb{R}$, we have $|\mathbb{N}| \leqslant |\mathbb{R}|$. We only need to show that $|\mathbb{N}| \neq |\mathbb{R}|$. Since the open interval $(0, 1)$ has the same size as $\mathbb{R}$, it is enough to show that $|\mathbb{N}| \neq |(0, 1)|$. Our proof is by contradiction.

Suppose that there is a one-to-one correspondence between $\mathbb{N}$ and $(0, 1)$, say

$$\begin{array}{rcccccc} \mathbb{N} = \{ & 1, & 2, & 3, & 4, & 5, & \dots \} \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ (0,1) = \{ & a_1, & a_2, & a_3, & a_4, & a_5, & \dots \} \end{array}$$

Let us write out the decimal expansion for each $a_i \in (0, 1)$ as

$$a_i = 0 \,.\, a_{i1}\, a_{i2}\, a_{i3}\, a_{i4}\, a_{i5} \,\cdots$$

where $a_{ij} \in \{0, 1, 2, \dots, 9\}$. Thus we have a perfect matching between $\mathbb{N}$ and $(0, 1)$ given by:

$$\begin{aligned} 1 &\leftrightarrow a_1 = 0 \,.\, a_{11}\, a_{12}\, a_{13}\, a_{14}\, a_{15} \,\cdots \\ 2 &\leftrightarrow a_2 = 0 \,.\, a_{21}\, a_{22}\, a_{23}\, a_{24}\, a_{25} \,\cdots \\ 3 &\leftrightarrow a_3 = 0 \,.\, a_{31}\, a_{32}\, a_{33}\, a_{34}\, a_{35} \,\cdots \\ 4 &\leftrightarrow a_4 = 0 \,.\, a_{41}\, a_{42}\, a_{43}\, a_{44}\, a_{45} \,\cdots \\ 5 &\leftrightarrow a_5 = 0 \,.\, a_{51}\, a_{52}\, a_{53}\, a_{54}\, a_{55} \,\cdots \end{aligned}$$

etc.

We will obtain a contradiction by finding a number $x \in (0,1)$ that is not in this list. Our value of $x$ will be described in terms of its decimal expansion

$$x = 0\,.\,x_1\,x_2\,x_3\,x_4\,x_5\,\cdots$$

where $x_i \in \{0,1,2,\ldots,9\}$. To ensure that $x \neq a_1$, simply choose the first digit $x_1 \neq a_{11}$; this leaves nine choices for the digit $x_1$. To ensure that $x \neq a_2$, simply choose the second digit $x_2 \neq a_{22}$. To ensure that $x \neq a_3$, simply choose the third digit $x_3 \neq a_{33}$. And so on. In general, we choose the $i$th digit of $x$ different from the $i$th digit of $a_i$, so $x \neq a_i$. Thus $x \in (0,1)$ does not appear in the list $\{a_1, a_2, a_3, a_4, a_5, \ldots\}$, contradicting our assumption that all elements of $(0,1)$ were in this list. $\square$

There is a small defect in the proof we have given, due to the fact some real numbers in $(0,1)$ have two different decimal expansions; for example $\frac{1}{2}$ is expressed in two ways as

$$0.500000000\ldots = 0.499999999\ldots.$$

This may be easily fixed: when choosing each digit $x_i$, simply avoid 0 or 9. Since we also want to choose $x_i \neq a_{ii}$, this still leaves either seven or eight possible choices for $x_i$, which is plenty. Of course one could choose to systematically define

$$x_i = \begin{cases} 3, & \text{if } a_{ii} \neq 3; \\ 7, & \text{if } a_{ii} = 3. \end{cases}$$

I previously noted that very few theorems about real numbers are proved using the decimal expansions. Cantor's proof, given above, is an important exception. The most important feature of this proof is the appearance of Cantor's *diagonal trick* which is a key ingredient in many other proofs in set theory, the theory of computation, point set topology, and mathematical logic.

Now let us consider names for numbers. What does '3' mean? How do we define 3? 'Threeness' is an abstraction which is concretely represented by any set with 3 objects. The set $U = \{\text{red, green, blue}\}$ will do the job quite as well as any other. We can't define 3 as 'the number of elements in any set with 3 objects' because such a definition would be circular. However, if everyone understands the principle of one-to-one correspondence (at least our poor shepherd does), then we know what $|S| = |U|$ means (namely, the elements of $S$ can be matched up with those of $U$) and we can write $|S| = 3$ as a shorthand in this case. In the same way, we would like names for the size of infinite sets. And the symbol '$\infty$' is not specific enough, since some infinite sets are larger than others. We traditionally write $|\mathbb{N}| = \aleph_0$ and we variously denote $|\mathbb{R}| = \mathfrak{c} = \beth_1 = 2^{\aleph_0}$. Here $\aleph$ ('Aleph') and $\beth$ ('Beth') are the first two letters of the Hebrew alphabet.

There is an infinite staircase of possible cardinalities of infinite sets, with $\aleph_0$ being the smallest. While the set $\mathbb{N}$ is infinite, it is 'just barely' infinite—it is as small as any infinite set can be. The set $\mathbb{R}$ is strictly larger. The set $F$ consisting of all functions $\mathbb{R} \to \mathbb{R}$ is even larger than $\mathbb{R}$. Whether or not there are any sets larger than $\mathbb{N}$ but smaller than $\mathbb{R}$ cannot be answered with the usual axioms of mathematics (the so-called ZFC axioms) and so sometimes an additional axiom is adopted (the *Continuum Hypothesis*) which asserts that there is no such set. At other times, mathematicians may choose to admit the existence of other sets $S$ with $|\mathbb{N}| < |S| < |\mathbb{R}|$. This issue is of very little relevance to the typical applications of mathematics.

However, the fact that $|\mathbb{N}| < |\mathbb{R}|$ is quite relevant. Every set of cardinality $\aleph_0$ is called *countable*; every set of cardinality greater than this is called *uncountable*. Thus $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$ and $\mathbb{Q}$ are countably infinite, whereas $\mathbb{R}$ is uncountable. The resolution of our 'paradox' of probabilities is that for the coin game (Scenario II), the set of possible outcomes was only countably infinite. If we understand the set of possible temperatures (Scenario III) as $\mathbb{R}$, or at least an interval in $\mathbb{R}$, then there are uncountably many possible outcomes. We can meaningfully talk about sums of countably many values. . . we do this all the time in Calculus II, and it is a big part of the syllabus of Math 2205. We cannot assign any meaning to the sum of uncountably many numbers, unless most of the numbers are zero, in which case it is really a sum of countably many numbers; and even then, it probably wouldn't mean what you want it to mean. In our probability example, the sum of uncountably many zeroes still doesn't equal 1.

It is best to think about countable sets as those sets whose elements can be enumerated in a (possibly infinite) list. A set of the form

$$S = \{a_1, a_2, a_3, \ldots\}$$

is countable because of the one-to-one correspondence $\mathbb{N} \leftrightarrow S$ given by $n \leftrightarrow a_n$. To say that $\mathbb{R}$ is uncountable means that it is to big a set to express in the form $\{a_1, a_2, a_3, \ldots\}$.

For any two sets $S$ and $T$, we may apply the cardinal rules (C1) and (C2) to compare their size: either they have the same size, or one is smaller than the other. But some choices of sets, this may be challenging. For example, consider the set $(0, 1] = \{x \in \mathbb{R} : 0 < x \leqslant 1\}$. Since

$$(0, 1) \subset (0, 1] \subset \mathbb{R},$$

we have

$$|(0, 1)| \leqslant |(0, 1]| \leqslant |\mathbb{R}|.$$

Since $|(0, 1)| = |\mathbb{R}|$, we should conclude that $|(0, 1]| = |\mathbb{R}|$. But this means that there is a perfect matching between $(0, 1]$ and $\mathbb{R}$. Can you find such a one-to-one correspondence? It is not easy! There is however a bijection $h : (0, 1] \to (0, 1)$ given by

$$h(x) = \begin{cases} \frac{1}{n+1}, & \text{if } x = \frac{1}{n} \text{ for some natural number } n; \\ x, & \text{otherwise.} \end{cases}$$

Thus $h(1) = \frac{1}{2}$, $h(0.7) = 0.7$, $h(\frac{1}{7}) = \frac{1}{8}$, $h(0) = 0$, etc. Composing $h$ with a bijection $f : (0,1) \to \mathbb{R}$ as given above, gives a bijection $f \circ h : (0,1] \to \mathbb{R}$. Now how about a bijection between $[0,1]$ and $\mathbb{R}$? Fortunately an explicit choice of bijection is not usually needed, and instead we freely use the following theorem of Cantor, Schröder and Bernstein:

**Theorem.** *If $|S| \leqslant |T|$ and $|T| \leqslant |S|$, then $|S| = |T|$. In other words, if there is a matching between $S$ and a subset of $T$, and there is a matching between $T$ and a subset of $S$, then there is a matching between $S$ and $T$.*

The proof is not terribly difficult, but we will omit it. It can be presented in about one lecture in a typical undergraduate class, but this is more time than we want to devote to the subject. However we feel it is important to at least state the result; otherwise it is not clear that it is reasonable to rank sets according to their size the way we would like.

Let us conclude with the observation that $|\mathbb{R}^2| = |\mathbb{R}|$. Here $\mathbb{R}^2 = \{(x,y) : x, y \in \mathbb{R}\}$ which may be identified as the set of points in the Euclidean plane. Since $|(0,1)| = |\mathbb{R}|$, it is enough to find a bijection between the set of points in the square $(0,1)^2 = \{(x,y) : 0 < x < 1,\ 0 < y < 1\}$ and the set of points in the open interval $(0,1)$. Such a correspondence is given by

$$(0\,.\,x_1\,x_2\,x_3\,x_4\,\ldots\,,\,0\,.\,y_1\,y_2\,y_3\,y_4\,\ldots) \;\leftrightarrow\; 0\,.\,x_1\,y_1\,x_2\,y_2\,x_3\,y_3\,x_4\,y_4\,\ldots\,.$$

Once again there is a slight defect in this correspondence, due to the fact that some real numbers have two different decimal expansions. This obstacle can be easily overcome using the Cantor-Schröder-Bernstein Theorem. A similar argument works for all $n \in \mathbb{N}$ to show that $|\mathbb{R}^n| = |\mathbb{R}|$. We see that there are just as many points in a single interval such as $(0,1)$, as there are in $\mathbb{R}^3$. Does this say that there are just as many points on a line segment as there are in the entire universe? It does, except for the fact that real numbers do not strictly represent points on a physical line, nor does every point in $\mathbb{R}^3$ strictly represent a point in physical space. But these matters are more metaphysical than mathematical...

The following history of the Bernstein-Cantor-Schröder Theorem, adapted from G. Moore's *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, Springer-Verlag, 1982, appears in *The Structure of Complete Designs*, Stuart A. Kurtz et. al., which appears as a chapter in Complexity Theory Retrospective, ed. Alan L. Selman, 1990.

**The Cantor-Bernstein Theorem** *If there is a 1-1 map from A into B and a 1-1 map from B into A, then there is a 1-1 correspondence between A and B.*

The standard proof of this theorem (see the proof of Theorem 1) is so simple and direct it is difficult to realize that a satisfactory proof of the theorem was hard to obtain. What follows is a partial history of the result taken from Moore's excellent book [Moo82].

**1882** Cantor claims (without proof) the theorem in a paper.

**1883** Cantor proves the theorem for subsets of $\mathbf{R}^n$ using the Continuum Hypothesis.

**1883** Cantor poses the general theorem as an open problem in a letter to Dedekind.

**1884** Cantor again claims (without proof) the theorem in a paper.

**1887** Dedekind proves the theorem in his notebook and promptly forgets about solving the problem. This particular solution is first published in 1932 in Dedekind's collected works.

**1895** Cantor states the theorem as an open problem in a paper.

**1896** Burali-Forti proves the theorem for countable sets.

**1897** Bernstein (a Cantor student) proves the general theorem using the Denumerability Assumption, a weak form of the axiom of choice. Cantor shows the proof to Borel.

**1898** Schröder publishes a "proof" of the theorem. Korselt points out to Schröder that the proof is wrong. Schröder's proof turns out to be unfixable.

**1898** Borel publishes Bernstein's proof in an appendix of Borel's 1898 book on set theory and complex functions.

**1899** Dedekind sends Cantor an elementary proof of the theorem, which from Moore's description sounds like the proof used today.

**1901** Bernstein's thesis appears (with his proof).

**1902** Korselt sends in a proof of the theorem (along the lines of Dedekind's proof) to *Mathematische Annalen.* Korselt's paper appears in 1911!

**1907** Jourdain points out that the use of the Denumerability Assumption in Bernstein's proof is removable.

Moore suspects that Cantor had a proof of the theorem that used the well ordering principle (which turns out to be equivalent to the axiom of choice). Cantor was unclear for many years whether the well ordering principle was "a law of thought" or something that needed proof. So, the fact that the theorem came and went several times might have been in part a function of Cantor's uncertainty about well ordering.