

# Numbers Can Be Fun

## An Imaginary Tale: The Story of the Square Root of $-1$

by Paul J. Nahin (Princeton University Press, 1998)

Review by Paul M. Conlin

If your mathematical education resembled that of many actuaries, you probably specialized in statistics. This most likely prevented you from allocating as much time as you may have otherwise liked to have spent on other subject areas, such as complex (or imaginary) numbers. And if you were like me, you may also have mentally classified complex numbers with activities such as proving that 1 plus 1 equals 2, and thus congratulated yourself on steering clear of an impractical area of study. Paul J. Nahin, an electrical engineering professor at the University of New Hampshire, has recently released a softcover edition of his book (originally published in 1998) on the history of complex numbers. Nahin demonstrates (for all who were not already aware) that complex numbers are richly practical in a number of real-world areas. (At the end of this review, I propose an actuarial topic which can be enlightened by consideration of complex numbers.)

Nahin begins by summarizing the efforts of Renaissance-era Italian mathematicians to solve the cubic equation,  $x^3 + a_1x^2 + a_2x + a_3 = 0$ . The solution to the quadratic equation,  $ax^2 + bx + c = 0$ , had been known since Babylonian times, but as recently as 1494, the solution of the cubic was declared impossible. Over the course of the following century, several Italian mathematicians solved the cubic equation. But in so doing, they stirred up a whole new series of questions to be pondered by future generations of mathematicians, such as William Stirling Hamilton, John Wallis, Caspar Wessel, Abraham de Moivre, Leonhard Euler, and Johann Carl Friedrich Gauss. This led to the current understanding of the square root of minus 1, complex numbers (any number  $a + bi$  having a real and imaginary component), and the geometry of the complex plane.

### Solution to the Cubic Equation

The solution to the cubic equation turns out to be a variation on the solution to the quadratic. The quadratic has either two real or two complex solutions, depending on whether  $b^2 - 4ac$  is positive or negative (since the quadratic equation involves taking the square root of that quantity). The cubic solution is similar, but with a twist. It turns out that a cubic equation has three solutions (consistent with the fundamental theorem of algebra, which states that an  $n$ -degree polynomial has  $n$  solutions). The three solutions can be either three real numbers or one real number and two complex numbers (geometrically, the curve representing the cubic passes thru the  $x$ -axis either one or three times, but not zero or two times). Further, if a given cubic has two complex solutions, then the two complex solutions will always be conjugate pairs ( $a+bi$ ,  $a-bi$ ). Conjugate pairs have the neat property of being complex numbers that generate a real number if they are summed  $\{a+bi + a-bi = 2a\}$  or multiplied  $\{a+bi * a-bi = a^2 + b^2$ , since  $i^2 = -1\}$ . This presence of conjugate pairs in the solution to the cubic is what facilitates the useful linear algebra result that a  $3 \times 3$  matrix of real numbers has eigenvalues (i) that sum to the sum of the diagonal elements of the matrix and (ii) the product of which equals the determinant of the matrix, even though two of the eigenvalues might be complex.

### Geometric Representation of $i$

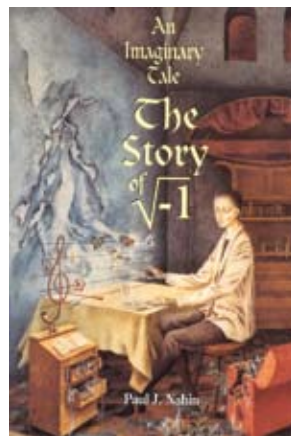
Just as an understanding of the cubic equation benefits from trying to understand both the algebra and the geometry of what

is occurring, Nahin also demonstrates that there is an actual geometrical way of understanding  $i$ . If you consider that  $i^0 = 1$ ,  $i^1 = i$ ,  $i^2 = -1$ ,  $i^3 = -i$ ,  $i^4 = 1$ ,  $i^5 = i$ ,  $i^6 = -1$ ,  $i^7 = -i$ , etc, then what is actually happening is that a unit vector is being rotated counterclockwise in a two-dimensional plane, where the horizontal axis has all the real numbers and the vertical axis has all the multiples  $bi$  of  $i$ . Multiplying a real number  $b$  times  $i$  rotates

the vector of that number 90 degrees in the complex plane. This breakthrough is what led to the contemporary custom of representing complex numbers  $a+bi$  as being located on a two-dimensional plane, with the  $a$  being graphed along the  $x$ -axis and the  $b$  being graphed along the vertical axis. Complex numbers cannot be ordered like real numbers, but they can be

added, subtracted, multiplied, and divided; they can occur in vectors that have an inner product; and they can occur in matrices that have eigenvalues and eigenvectors. We can even raise a number to an imaginary power (did you know that  $i^i = .2078.....$  ?).

Once this rotational property of  $i$  was understood, de Moivre could demonstrate his important trigonometric identity  $\{\cos a + i \sin a\}^m = \cos (a/m) + i \sin (a/m)$  (again, picture the complex plan, with the  $\sin$  element being rotated onto the vertical axis by means of multiplying it by  $i$ ), and Euler could derive his famous equation  $e^{ia} = \cos a + i \sin a$ , the special case of which (when  $a = \pi$ ) is what Richard Feynman calls "the most remarkable formula in math,"  $e^{i(\pi)} + 1 = 0$ . This formula has the five most important numbers in math and the four most important operations (addition, multiplication, exponentiation, and equality), each of which occur once and only once in the formula.



### Fascinating, but so What?

Nahin's book is particularly good at staying on message—complex numbers have fascinating properties and generate fascinating results but are also relentlessly useful. He gives many interesting examples, including the famous electrical engineering phase-shift oscillator circuit, which is the basis of all modern semiconductors.

He also gives an example that I would propose has implications for actuaries. One of the frustrating trends of the past 15 years or so has been the tendency of financial institutions (Enron, Mutual Benefit Life Insurance Co., General American Life Insurance Co., etc.) to not only fail to diversify risk but, in fact, to multiply risk (see Michael Mateja's discussion in the general session of the Society of Actuaries 2004 Anaheim, Calif., spring meeting, Vol. 18 of the Record of the Society of Actuaries). This violates an important property of real numbers, which is so fundamental that it's covered in most real analysis courses as the Cauchy-Schwartz Inequality. The insurance analogue of this is used in risk-based capital (RBC) calculations, where the total required surplus of an insurance entity is the square root of  $\{(C1+C3)^2 + C2^2 + C4^2\}$ , where C1 is the total required surplus due to C1 risk, etc. The insurance entity is presumed to benefit from diversification by the resultant RBC being less than the sum of the  $C1 + C2 + C3 + C4$ .

Nahin shows that the Cauchy-Schwartz Inequality doesn't hold for complex numbers. In other words, in a world of complex numbers, the total risk is potentially greater than the sum of the risks. Do financial institutions operate in a world of complex numbers? I don't know. But complex numbers allow us to perceive a way in which they may multiply risk, not diversify them. ●

PAUL M. CONLIN is a fellow of the Society of Actuaries and an actuary with Aetna in Hartford, Conn.

## Data Quality and Record Linkage Techniques

By Thomas N. Herzog, Fritz Scheuren, and William E. Winkler  
(Springer Science and Business Media, 2007)

Review by Bruce D. Schobel

Tom Herzog has a history of writing books (with various co-authors) about relatively mundane topics that most mathematically literate people believe they already understand pretty well—until they read the book.

His 2002 book with Graham Lord, *Applications of Monte Carlo Methods to Finance and Insurance* (see my review in the September/October 2004 issue of *Contingencies*), opened many new windows onto methods that actuaries and others use with great frequency without having a deep understanding of how they work—or of the pitfalls that await the unwary. That book taught me more about random-number generators than I thought could be known! I have had this experience reading Tom's work often enough for it to be familiar, but it's nevertheless disconcerting.

Herzog, chief actuary at the U.S. Department of Housing and Urban Development, wrote his latest book with Fritz Scheuren of the National Opinion Research Center and William Winkler of the U.S. Census Bureau. All three authors have doctorates in probability, statistics, or related fields and work routinely with huge databases that contain various kinds of errors and duplications. Improving the quality of those databases is the focus of their concern in this interesting book.

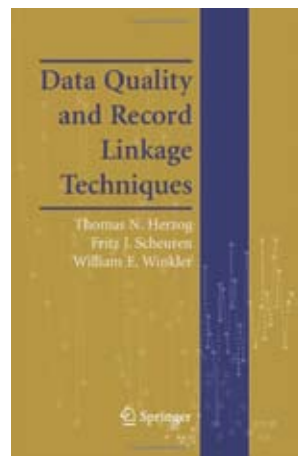
The book has four major sections. The first defines data quality, which in a sense means identifying the various kinds of errors that can exist in a database. Each record in a database may contain numerous elements, any of which may be inaccurate

in one way or another. Some errors may be irrelevant, depending on how the database is being used. In using a Census Bureau database, for instance, a demographer may care a lot about the accuracy of birthdates

but little about addresses. On the other hand, a legislative assistant who is attempting to redraw the boundaries of congressional districts cares very much about the accuracy of address information but little about age. All researchers care about eliminating duplicate records, which inevitably occur. People change their addresses, for instance, and appear for a while in databases at both the old

and new locations. Duplications also occur when databases are merged but identical records are not recognized as being identical. The authors provide many real-life examples and describe tests of data quality that can be applied.

The second section of the book delves more deeply into specialized tools that can be used to improve the quality of databases—or at least to estimate the number of errors. One such tool is the capture-recapture procedure. The authors provide many complex examples of the technique but also give a simple example that outlines how it works. If you wish to estimate the number of fish in a pond, you could collect a sample of fish one day, tag them and release them back into the pond, then take another sample the next day, counting how many in the second sample had been tagged the first day. Using capture-recapture methods, the total number of fish in the pond could be estimated. The authors



also discuss the costs associated with data editing and when those costs outweigh the value. Finally, this section contains several chapters focused on record linkage—how to recognize that two seemingly separate records really relate to the same item. One record may have the name Bob Smith and the other Robert Smith, but if the addresses are the same, then we might assume that the records are really one. Recognizing this connection is simple for humans, who do this almost instinctively, but developing techniques for automated record linkage is much more difficult.

The book's third section contains a series of case studies. The most interesting to me was the chapter on multiple Social Security number issuance and identification of deaths. For years, there were few safeguards to prevent people from applying for multiple Social Security numbers. Some people just filed a new application

any time anybody (employers, banks, etc.) asked them for their Social Security number. Most of the time, the Social Security Administration (SSA) would recognize that the applicant already had a number and would issue a new card with the already issued number, but sometimes a new number would be issued. The techniques in this book show how the extent of that problem could be accurately estimated. The Social Security chapter also discusses how beneficiaries are linked to death records. All of us have read newspaper stories about perfectly healthy people being declared dead by SSA. Obviously, those are a consequence of inaccurate data matches.

The last part of the book briefly touches on other related topics, one of which is maintaining confidentiality in micro-data. As data sets become smaller, identification of individual records becomes more likely, and this can be a problem, especially

when confidentiality must be protected (when dealing with medical records, for example). The book explains how records can be masked. The simplest method is elimination of certain data elements. For instance, an income field can be dropped if such information is considered too sensitive to risk identification of individuals in a database. But elimination of data can make a database unusable. Other techniques, such as swapping data elements or rounding data to obscure identifying characteristics, are more effective.

This book is definitely not light reading, but it's interesting and informative. Anyone who works with large databases should read it.

**BRUCE D. SCHOBEL** is president of the Society of Actuaries, a member of the *Contingencies* Editorial Advisory Board, and vice president and actuary for New York Life Insurance Co. in New York.

## Does your lawyer understand:

- The difference between Schedule P and a Schedule B?
- IBNR estimates based on the BF method?
- Reinsurance commutations?
- Viatical settlements?
- The ASOPs?

*We do.*

*We can resolve your complex actuarial, pension and insurance disputes.*

To discuss your dispute, contact Reinhart lawyers Tim Muth or Sandy Brown at 800-553-6215

Milwaukee  
414-298-1000  
800-553-6215

reinhartlaw.com

**Reinhart**  
Boerner Van Deuren s.c. Attorneys at Law