Mathematics by experiment: plausible reasoning in the 21st century

J. Borwein and D. Bailey
AK Peters Wellesley MA 2003
ISBN 1-56881-211-6

Experimentation in mathematics: computational paths to discovery

J. Borwein, D. Bailey and R. Girgensohn
AK Peters Wellesley MA 2004
ISBN 1-56881-136-5

Perhaps the first piece of advice I was given by my PhD supervisor about doing mathematics was

> "*It is always easier to prove something when you know it is true.*"

For me, this statement highlights the difference between the way that most of us *do* mathematics and the way that we *present* it.

When we publish a proof we are often like a magician showing their latest trick — a slick, polished and beautiful performance, that (hopefully) entertains the audience and leaves them in awe of the spectacle and impressed by our ingenuity. We move in a crescendo of logic from definitions to lemmas to our main theorem; always onwards and upwards.

On the other hand, we do not actually *do* mathematics in this way. In fact it is probably more accurate to say that we start with the theorem and work backwards:

> "*I have the result, but I do not yet know how to get it.*"
>
> — C. F. Gauss

All mathematicians build their intuition about problems and objects by looking at examples. When this intuition becomes strong enough, we *know* the result before we have a proof of it. And once we have the theorem we set about proving it. Arguably mathematics has been this way for a very long time; while it is different from other areas of knowledge in that it has the certainty of proof, as practitioners we are not immune from getting our hands a little dirty with some experimentation — though we are generally not willing to admit it and tend to hide it if possible.

The message of these two books is that it is time to embrace experimentation as part of mathematics, not hide it. And this is now possible because the computer has made wide-spread, systematic experimentation a reality.

Moore's law, and all the engineers who have perpetuated it, have made computers powerful, cheap and ubiquitous. We now have an abundance of raw computing power and sophisticated mathematical software, both free (such as GAP) and commercial (such as Maple and Mathematica), at their disposal. This gives us a "laboratory" in which we can perform numerical and symbolic explorations on a grand scale without pages of painful by-hand calculations. In so doing we can test conjectures and guess entirely new ones. This approach is leading to new results discovered and even proved partially or entirely with the aid of computers.

These two books are devoted to exploring this approach and are aimed at a wide cross-section of mathematically-trained readers from (roughly) honours level and up.

The first, *Mathematics by experiment*, describes the ideas behind experimental mathematics (giving those words a far more concrete definition than I have done here) and devotes some time to the more philosophical implications suggested by the title — including a very interesting section on paradigm shifts. The bulk of the book is made up of a series of examples of experimental mathematics in action. Reflecting the expertise of the authors, the examples come mainly from number theory and combinatorics, covering topics such as the digits of π and normality, but also included are smaller examples from a wide range of other topics such as chaos theory and knot theory. The book is perhaps best read with a computer nearby so that you can tinker and try things out as you read. The examples are also complemented by commentaries and exercises for the reader.

The chapter on π gives an account of various computational methods that have been employed to find more and more terms of its expansion in different bases. This culminates in the famous result by Peter Borwein, David Bailey and Simon Plouffe:

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i}\left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6}\right).$$

The importance of this result lies in that it can be used to compute individual digits of π — one can compute the d^{th} binary digit without having to compute any of the previous digits! The formula was not discovered by formal reasoning or symbolic calculations, but rather was discovered by high precision *numerical* calculations.

The computation of digits of π then leads naturally into a more general discussion of the *normality* of numbers — a number is *normal* if its expansion (in a given base) behaves like a random sequence. While it has been shown that almost all numbers are normal, the only numbers *proved* to be normal are very contrived. On the other hand, numerical analyses show many common mathematical constants, such as π and $\sqrt{2}$, *appear* to be normal. The authors describe how results like that given for π above, have lead to significant progress in the theory of normal numbers. A careful experimental "verification" of the normality of algebraic numbers is described later in the book.

The second-last chapter of *Mathematics by experiment* describes many of the numerical tools and techniques that are required (such as high-precision arithmetic and integer relation detection) and also give many useful links to implementations of these methods. The book then ends with a reprint of the article "*Making sense of experimental mathematics*" by J.M. Borwein, P.B. Borwein, R. Girgensohn and S. Parnes.

The second text, *Experimentation in mathematics* continues with further chapters of examples and numerical and symbolic techniques. This text moves through a broader range of topics and is more demanding of the reader than the first. Number theory and combinatorics are prominent amongst the topics covered; the chapter on zeta functions details some of the work by the authors on special values of zeta and multizeta functions and shows how this work grew from and inspired their interest in experimental methods. Again, while topics closest to the authors' experience, form the bulk of the book, there are many sections that cover problems from other areas of mathematics including probability theory and Fourier series.

While the both books may be read and used independently, I would suggest reading the *Mathematics by experiment* (particularly because of the philosophical and historical discussions) before delving into the heavier *Experimentation in mathematics*.

In addition to a large quantity of mathematical theory, examples and problems they contain, the books are littered with numerous interesting (and often colourful) stories and histories of people and theorems.

To summarise, I do not think I that have had the good fortune to read two more entertaining and informative mathematics texts.

Andrew Rechnitzer
Department of Mathematics, University of Melbourne, VIC 3010
E-mail: A.Rechnitzer@ms.unimelb.edu.au

◇ ◇ ◇ ◇ ◇ ◇

Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond

B. Schölkopf and A.J. Smola
MIT Press 2002
ISBN 0-262-19475-9

According to Tomaso Poggio and Steve Smale [6]:
"We believe that a set of techniques based on a new area of science and engineering becoming known as "supervised learning" will become a key technology to extract information from the ocean of bits around us and to make sense of it."
When mathematical scientists of this stature make such strong statements, we should take note. In broad terms, learning theory (also known as machine learning) deals with developing rules for making decisions based on learning from examples or data. For example, we may develop rules for classifying a new email message as junk mail by using the experience of previous messages.

In recent years there have been several books written on learning theory that may appeal to readers of the *Gazette*: see the references below for a selection. Some, such as Hastie *et al.* [3], may appeal to statisticians; others, such as those by Cristianini and Shawe-Taylor [2], [8], or Kecman [4] may appeal to those with a computer science bent. The present book under review by Bernhard Schölkopf (Max Planck Institute for Biological Cybernetics) and Alexander Smola (ANU) may appeal to mathematicians. Learning theory brings together many aspects of pure mathematics, mathematical modelling, computational mathematics, probability, statistics, and computer science. Reading through books on the subject makes one wonder about the wisdom of the ways in which universities compartmentalise these parts of mathematics in the curriculum or in different departments.

Perhaps many of us associate the phrase "machine learning" with neural networks and probably don't think more about it. Schölkopf and Smola have written a book of 626 pages on machine learning and rarely mention neural networks. They have produced an introduction to kernel based methods of machine learning which use ideas from classical machine learning theory, optimization, and mathematical analysis—including the ubiquitous reproducing kernel Hilbert spaces. The canonical example of this approach is the Support Vector Machine (SVM) about which they write:
"...successful applications have demonstrated that SVMs not only have a more solid foundation than artificial neural networks, but are able to serve as a replacement for neural networks that perform as well or better, in a wide variety of fields."
(For an extensive discussion of the links between approximation theory and neural networks, see [5].)

The book opens with a tutorial. Chapter 1 is an informal introduction to SVMs and associated ideas intended to orient the reader. The authors use a simple example of classifying certain objects into one of two classes (e.g. classify an email message as junk or not junk). A subset of the objects is used for training, and each of these objects is mapped into an inner product space known as the feature space. Classification is then a process of determining a hyperplane which separates the vectors representing the objects.