Experimental Mathematics and Computational Statistics

David H. Bailey [*]   Jonathan M. Borwein [†]

July 6, 2009

Abstract

The field of statistics has long been noted for techniques to detect patterns and regularities in numerical data. In this article we explore connections between statistics and the emerging field of "experimental mathematics." These includes both applications of experimental mathematics in statistics, as well as statistical methods applied to computational mathematics.

1   Introduction

All truths are easy to understand once they are discovered; the point is to discover them. — attributed to Galileo Galilei

Knowing things is very 20th century. You just need to be able to find things. — Danny Hillis

In a provocative recent article entitled "The End of Theory," Chris Anderson, the Editor-in-Chief of *Wired*, heralds a new mode of scientific inquiry where exploding repositories of data, analyzed using advanced mathematical and statistical techniques in the same manner as Google has analyzed the Internet, are sufficient to render the traditional scientific method (hypothesize, model, test) obsolete: "The new availability of huge amounts of data, along with the statistical tools to crunch

these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?" [2].

Kevin Kelly, in response to Anderson's article, makes a more modest statement: "My guess is that this emerging method will be one additional tool in the evolution of the scientific method. It will not replace any current methods (sorry, no end of science!) but will complement established theory-driven science. ... The model may be beyond the perception and understanding of the creators of the system, and since it works it is not worth trying to uncover it. But it may still be there. It just operates at a level we don't have access to." [28].

The present authors do not see that traditional model-based science is in any danger of obsolescence, but we do agree that the explosion of data and novel computational techniques offer routes to scientific truth that are of a qualitatively different nature than in previous eras—whether or not a model lurks behind the curtain. At the least, it behooves us to explore these methods and see how they can complement traditional approaches.

In this article, we offer some examples where aggressively utilizing of modern computational tools have yielded results in an arena that is not often thought of as amenable to data-driven analysis, namely mathematical research. We will demonstrate that this new methodology has numerous connections with computational statistics in both directions—the experimental math methodology can be productively applied to certain problems in statistics, and statistical methods can be applied in experimental math.

2 A Brief Overview of Experimental Mathematics

"Experimental mathematics" is a name that has been loosely given to a new mode of doing mathematical research where the computer is used as a "laboratory," and the "data" are the results of mathematical computation. Using this methodology, we can "see" results long before we can rigorously prove them, and in fact, the experimental results may point the direction of formal proofs. Some techniques utilized in this methodology include high-precision numerical computation, integer relation detection, symbolic computation, statistical analysis and visualization tools. These methods are described by ourselves and others elsewhere, e.g., [19, 20], but we will briefly mention three particularly important numerical methods here:

2.1 High-Precision Arithmetic

High-precision arithmetic is arguably the most heavily used single technique in experimental mathematics—digit expansions are in many cases the principal "raw

data" of analysis. The main driver behind this requirement is the fact that integer detection methods (see next subsection), which are used to discover new mathematical identities, require very high precision numerical inputs to obtain numerically meaningful results. Many problems require several hundred digits of precision; one application required 50,000-digit precision. From a programming point of view, computer implementations of high-precision arithmetic typically represent a high-precision datum as a string of computer words, where the first few words give the sign, exponent and array size, and the remaining words contain successive sections of the binary mantissa. A brief survey of techniques used to perform arithmetic operations on these data structures is given in [19, pg 218–229]. Much relies on implementation of a good complex *fast Fourier transform* to perform multiplication efficiently. Software is widely available for performing such computation, for example the "QD" package, which performs double-double and quad-double precision arithmetic, and the "ARPREC" package, which performs arbitrary precision arithmetic. These two packages permit researchers to convert existing C/C++ and Fortran-90 programs, with only a few minor changes, to use high-precision arithmetic. They are available from

<div align="center">

`http://crd.lbl.gov/~dhbailey/mpdist` .

</div>

2.2 Integer Relation Detection

By "integer relation detection," we mean the following: Given n real numbers $(x_k, \ 1 \leq k \leq n)$ (typically presented as a list of real numbers given to high precision), determine if there are integers $(a_k, \ 1 \leq k \leq n)$ such that the inner product $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$ (or zero to within some specified precision). Such computations are typically used to see if a given computed value is given by a simple mathematical formula involving known constants. This is done by computing the given constant and all terms of the conjectured formula to high precision, and then using an integer relation algorithm to search for a relation. If a relation is found, then by solving the relation for the given constant, one obtains a formula. These computations do not constitute a rigorous proof that the formula holds, but with strong evidence of this sort it is often easier to find a proof. We will present several examples of this methodology below.

At the present time, the most widely used integer relation algorithm is the "PSLQ" algorithm of mathematician-sculptor Helaman Ferguson. A detailed statement of the PSLQ algorithm, together with some recently discovered "multi-level" and "multi-pair" variants of PSLQ, which run significantly faster and are better suited for parallel processing, are described in [19, pg. 230–234] and [12]. In normal operation of PSLQ on an input vector (x_k), the entries of the reduced vector gradually decrease in size until at the point of detection, the smallest entry abruptly drops to approximately the "epsilon" of the high-precision arithmetic being used. This behavior is shown in Figure 1, where the size drops by 180 orders of magnitude (to roughly 10^{-250}, using 250-digit arithmetic) when the underlying relation is discovered. The size of this abrupt drop can be taken as a "confidence ratio" in the
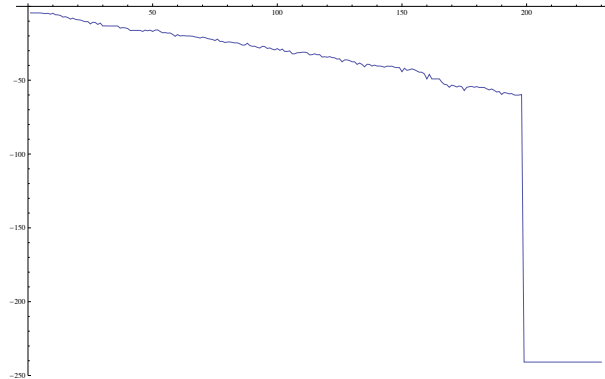
Figure 1: Plot of $\log_{10} \min |x_i|$ as a function of PSLQ iteration number

detection, in much the same way as one interprets confidence ratios in conventional statistical inference. A confidence ratio of 10^{20} or more almost always indicates a genuine underlying mathematical relationship rather than a spurious artifact of numerical round-off error.

It can be shown that if one wishes to find coefficients (a_k) of an integer relation for an input vector (x_k), where the maximum absolute value of x_k is roughly 10^d, then all the input values (x_k) must be specified to at least nd-digit precision, and key parts of the PSLQ algorithm itself must be performed to at least nd-digit precision (and usually a bit more), or else the true relation will be lost in a sea of spurious numerical artifacts. Informally speaking, this fact is based on a "conservation of information" argument that will be familiar to applied statisticians—one cannot extract more information from results than exists in the input.

The PSLQ algorithm can also provide "exclusion bounds," which sometimes are more valuable than detection results. For instance, it is not known whether or not *Catalan's constant* $G := \sum_{n \geq 0} (-1)^k / (2k+1)^2$ is the root of an algebraic equation with integer coefficients. PSLQ cannot answer this question definitively, but a 10,000-digit computation using PSLQ by the authors demonstrated that if G does satisfy some algebraic equation of degree 40 or less with integer coefficients, then the Euclidean norm $\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$ of these coefficients must exceed 5.58×10^{244}.

2.3 Numerical Integration

One other computational technique that we will briefly mention here, because it has found numerous recent applications (including several that are relevant to computational statistics) is high-precision *quadrature* (i.e., numerical integration). While a number of quadrature techniques, including such venerable methods as Gaussian quadrature, have been applied by us and others in experimental math studies, we have found that the little-known *tanh-sinh* scheme of Mori and Takahashi possesses several very attractive features, including: (a) it produces high-precision numerical values even for many integrand functions with singularities and/or infinite deriva-

4

tives at endpoints, (b) the cost of computing the requisite abscissas and weights scales only linearly with the number of evaluation points, whereas in Gaussian quadrature and most other schemes the cost increases at least as n^2; and (c) for many integrand functions of interest, doubling the number of evaluation points yields roughly double the number of correct digits in the result.

The tanh-sinh quadrature scheme transforms the integral of a given function $f(x)$ on the interval $[-1, 1]$ to an integral on $(-\infty, \infty)$ using the change of variable $x = g(t)$, where $g(t) = \tanh(\pi/2 \cdot \sinh t)$. Note that $g(x)$ has the property that $g(x) \to 1$ as $x \to \infty$ and $g(x) \to -1$ as $x \to -\infty$, and all derivatives rapidly approach zero for large positive and negative arguments. Thus one can write, for $h > 0$,

$$\int_{-1}^{1} f(x)\,dx \;=\; \int_{-\infty}^{\infty} f(g(t))g'(t)\,dt \;=\; h \sum_{j=-N}^{N} w_j f(x_j) + E(h), \qquad (1)$$

where $x_j = g(hj)$ and $w_j = g'(hj)$, and where N is chosen large enough that $|w_j f(x_j)| < \epsilon$ for $|j| > N$. Here $\epsilon = 10^{-p}$, where p is the numeric precision level in digits. Because $g'(t)$ and higher derivatives tend to zero very rapidly for large t, the resulting integrand $f(g(t))g'(t)$ typically is a smooth bell-shaped function, even if $f(t)$ itself is not well-behaved at the endpoints. For such bell-shaped integrands on an infinite interval, the Euler-Maclaurin formula implies that the error $E(h)$ in (1) decreases faster than any power of h. Additional details are given in [15] and [20, pg. 312–313].

Integrals in more than one dimension are more challenging. Even low precision computation of anything more than a handful of digits say by *quasi Monte Carlo* methods (qMC) is problematic. Both Gaussian quadrature and tanh-sinh quadrature can be extended to multiple dimensions in a straightforward manner, although the computational cost is greatly increased. For example, if 1,000 function evaluations are required to produce a result good to say 500 digits for a certain class of functions, then approximately 1,000,000 may be required in two dimensions, and 1,000,000,000 in three dimensions. Fortunately, modern highly parallel computing technology can produce results in reasonable elapsed time, provided that issues such as load balancing are handled properly. Some details of highly parallel implementations of high-precision numerical integration are given in [8].

In one recent application of these methods, the authors addressed *Problem 11275* in the February 2007 issue of the MAA MONTHLY [17], which problem asks to evaluate the iterated integral

$$Q := \int_0^{\infty} \int_y^{\infty} \frac{(x-y)^2 \log((x+y)/(x-y))}{xy \sinh(x+y)} \, dx \, dy.$$

By using a two-dimensional version of the tanh-sinh quadrature algorithm, we were able to calculate

$$Q \;=\; 1.15326598908047301786027529310599388545112440092244354 25100\ldots$$

5

Once we had obtained this numerical value, we used it as input to the Inverse Symbolic Calculator (ISC) tool, which is available at `http://ddrive.cs.dal.ca/~isc`. This tool attempts to recognize numerical constants by a combination of methods, including a table look-up procedure and an integer relation detection facility. At the time, it was not able to recognize this constant, but it was able to recognize

$$4Q/\pi^2 = 0.46740110027233965470862274996903778382842485181020\ldots$$
$$= \pi^2/4 - 2,$$

so that $Q = \pi^4/16 - \pi^2/2$. Once we "knew" the answer, it was not too difficult to find a rigorous proof [7]. Subsequently, improvements were made to the ISC tool, and now it is able to recognize the original constant.

3 Expectation Integrals

The best way to really explain what is meant by "experimental mathematics" is to present a few examples of this methodology in action. To that end, we offer the following examples, which are chosen, from among many examples we have seen, because of connections to statistics, probability theory or statistical mechanics of physics.

3.1 Box Expectation Integrals

Recently the present authors, together with Richard Crandall, used an experimental approach to study certain integrals that arise naturally as expectations of distance functions over random points in n-dimensional geometry. In particular, given parameter s and n-dimensional vector $q = (q_1, q_2, \cdots a_n)$, a *box expectation integral* is defined here as the expectation of the norm $|r - q|^s$ for points $r = (r_1, r_2, \cdots r_n)$ chosen in equidistributed random fashion over the unit n-cube:

$$X_n(s, q) := \langle |r - q|^s \rangle_{r \in [0,1]^n}$$
$$= \int_0^1 \cdots \int_0^1 \left((r_1 - q_1)^2 + \ldots + (r_n - q_n)^2 \right)^{s/2} dr_1 \cdots dr_n. \quad (2)$$

Two important variants are defined as:

$$B_n(s) := \int_0^1 \cdots \int_0^1 \left(r_1^2 + \ldots + r_n^2 \right)^{s/2} dr_1 \cdots dr_n, \quad (3)$$

$$\Delta_n(s) := \int_0^1 \cdots \int_0^1 \left((r_1 - q_1)^2 + \ldots + (r_n - q_n)^2 \right)^{s/2} dr_1 \cdots dr_n \, dq_1 \cdots dq_n.$$

In other words, $B_n(1)$ is the *expected distance of a random point from the origin* (i.e., the vertex) of the n-cube, while $\Delta_n(1)$ is the *expected distance between two random points* of the n-cube.

These integrals can be analyzed in much the same way as the *Monthly* integral of the previous section, namely by computing numerical values to high precision, and then using a tool such as the ISC to recognize the numerical values. The principal challenge here is the cost of computing multi-dimensional integrals to high precision, which cost increases very rapidly with dimension as noted earlier. However, as it turns out, in this case we were able to reduce the B and Δ integrals to one-dimensional integrals, dramatically reducing the computational cost:

$$B_n(s) \;=\; \frac{s}{\Gamma(1-s/2)} \int_0^\infty \frac{du}{u^{s+1}}\,(1-b(u)^n)\,,$$

$$\Delta_n(s) \;=\; \frac{s}{\Gamma(1-s/2)} \int_0^\infty \frac{du}{u^{s+1}}\,(1-e(u)^n)\,,$$

where

$$b(u) \;=\; \sum_{k=0}^\infty \frac{(-1)^k u^{2k}}{k!(2k+1)} \;=\; e^{-u^2}\sum_{k=0}^\infty \frac{2^k u^{2k}}{(2k+1)!!}\,,$$

$$e(u) \;=\; \sum_{k=0}^\infty \frac{(-1)^k u^{2k}}{(k+1)!(2k+1)} \;=\; e^{-u^2}\sum_{k=0}^\infty u^{2k}\left(\frac{2^{k+1}}{(2k+1)!!} - \frac{1}{(k+1)!}\right).$$

Here the !! notation means $n(n-2)(n-4)\cdots 2$ if n is even and $n(n-2)(n-4)\cdots 1$ if n is odd.

Based on numerical values that we obtained in this fashion, combined with mathematical analysis and some symbolic computation, we ultimately were able to find reasonable analytic evaluations in many cases. For example,

$$B_2(-1) \;=\; \log\left(3+2\sqrt{2}\right), \qquad B_3(-1) = -\frac{\pi}{4} - \frac{1}{2}\log 2 + \log(5+3\sqrt{3}),$$

$$B_1(1) \;=\; \frac{1}{2}, \qquad B_2(1) = \frac{\sqrt{2}}{3} + \frac{1}{3}\log\left(\sqrt{2}+1\right),$$

$$B_3(1) \;=\; \frac{\sqrt{3}}{4} + \frac{1}{2}\log\left(2+\sqrt{3}\right) - \frac{\pi}{24},$$

$$B_4(1) \;=\; \frac{2}{5} + \frac{7\pi}{20}\sqrt{2} - \frac{\pi}{20}\log\left(1+\sqrt{2}\right) + \log(3) - \frac{7}{5}\sqrt{2}\arctan\left(\sqrt{2}\right)$$
$$\qquad + \frac{1}{10}\mathcal{K}_0,$$

where the one unresolved term, namely

$$\mathcal{K}_0 \;=\; 2\int_0^1 \frac{\operatorname{arctanh}\left(\frac{1}{\sqrt{3+y^2}}\right)}{1+y^2}\,dy,$$

can be evaluated reasonably rapidly via the two-dimensional sum

$$\mathcal{K}_0 \;=\; \sum_{m,k\geq 0} \frac{2^{k+1}}{2m+1}\frac{I_{m+k}}{3^{m+k+1}} \;=\; \frac{2}{3}\sum_{p=0}^\infty I_p\left(\frac{2}{3}\right)^p \sum_{n=0}^p \frac{1}{2^n\,(2n+1)}.$$

Here $I_0 := 1/2$ and $I_m = (2mI_{m-1} + (3/4)^m/2)/(2m+1)$.

In a similar way, we were able to find the following evaluations of the Δ integrals.[1]

$$\Delta_2(-1) = \frac{2}{3} - \frac{4}{3}\sqrt{2} + 4\log(1+\sqrt{2}), \qquad \Delta_1(1) = \frac{1}{3},$$

$$\Delta_2(1) = \frac{1}{15}\left(2 + \sqrt{2} + 5\log(1+\sqrt{2})\right),$$

$$\Delta_3(1) = \frac{4}{105} + \frac{17}{105}\sqrt{2} - \frac{2}{35}\sqrt{3} + \frac{1}{5}\log(1+\sqrt{2}) + \frac{2}{5}\log(2+\sqrt{3}) - \frac{\pi}{15},$$

$$\begin{aligned}
\Delta_4(1) = {}& \frac{26}{15}\,\mathrm{G} - \frac{34\pi}{105}\,\sqrt{2} - \frac{16\pi}{315} + \frac{197}{420}\,\log(3) + \frac{52}{105}\,\log\left(2+\sqrt{3}\right) \\
& + \frac{1}{14}\,\log\left(1+\sqrt{2}\right) + \frac{8}{105}\,\sqrt{3} + \frac{73}{630}\,\sqrt{2} - \frac{23}{135} \\
& + \frac{136}{105}\,\sqrt{2}\arctan\left(\frac{1}{\sqrt{2}}\right) - \frac{\pi}{5}\,\log\left(1+\sqrt{2}\right) + \frac{4}{5}\,\alpha\log\left(1+\sqrt{2}\right) \\
& - \frac{4}{5}\,\mathrm{Cl}_2(\alpha) + \frac{4}{5}\,\mathrm{Cl}_2\left(\alpha + \frac{\pi}{2}\right),
\end{aligned}$$

$$\begin{aligned}
\Delta_5(1) = {}& \frac{65}{42}\,\mathrm{G} - \frac{380}{6237}\,\sqrt{5} + \frac{568}{3465}\,\sqrt{3} - \frac{4\pi}{189} - \frac{449}{3465} - \frac{73}{63}\,\sqrt{2}\arctan\left(\frac{\sqrt{2}}{4}\right) \\
& - \frac{184}{189}\,\log(2) + \frac{64}{189}\,\log\left(\sqrt{5}+1\right) + \frac{1}{54}\,\log\left(1+\sqrt{2}\right) \\
& + \frac{40}{63}\,\log\left(\sqrt{2}+\sqrt{6}\right) - \frac{5\pi}{28}\,\log\left(1+\sqrt{2}\right) + \frac{52\pi}{63}\,\log(2) \\
& + \frac{295}{252}\,\log(3) + \frac{4}{315}\,\pi^2 + \frac{3239}{62370}\,\sqrt{2} - \frac{8}{21}\sqrt{3}\arctan\left(\frac{1}{\sqrt{15}}\right) \\
& - \frac{52\pi}{63}\,\log\left(\sqrt{2}+\sqrt{6}\right) + \frac{5}{7}\alpha\log\left(1+\sqrt{2}\right) - \frac{5}{7}\,\mathrm{Cl}_2(\alpha) \\
& \frac{5}{7}\,\mathrm{Cl}_2\left(\alpha + \frac{\pi}{2}\right) + \frac{52}{63}\,\mathcal{K}_1,
\end{aligned}$$

where the unresolved quantity this time is the integral

$$\mathcal{K}_1 := \int_3^4 \frac{\mathrm{arcsec}(x)}{\sqrt{x^2 - 4x + 3}}\,dx.$$

In the above, $\alpha := \arcsin\left(2/3 - 1/6\sqrt{2}\right)$, $\mathrm{G} := \sum_{n\geq 0}(-1)^k/(2k+1)^2$ is the *Catalan constant*, Cl_2 is the order-2 *Clausen function*

$$\mathrm{Cl}_2(x) := \begin{cases} \sum_{k=1}^{\infty}\sin(kx)/k^n & \text{if } n \text{ even} \\ \sum_{k=1}^{\infty}\cos(kx)/k^n & \text{if } n \text{ odd}, \end{cases}$$

and $\mathrm{Li}_n = \sum_{k=1}^{\infty} z^k/k^n$ is the *polylogarithm* function of order n. Numerous other specific and general results have been obtained. See [10] for full details.

---

[1]In [10] the formulas for Δ_4 and Δ_5 both had the wrong sign in the coefficient of $\mathrm{Cl}_2(\alpha - \pi/2)$. Fortunately, PSLQ makes it very easy to error-correct such things as long as the basis elements are correct.

3.2   Ising Expectation integrals

Integrals very similar to the box integrals arise within many parts of mathematical physics; e.g., from the *Ising model* within statistical mechanics. Like the box expectation integrals, these integrals are cast as multi-dimensional expectations of certain simple algebraic functions. In recent studies with Crandall, we considered these classes:

$$C_n := \frac{4}{n!} \int_0^\infty \cdots \int_0^\infty \frac{1}{\left(\sum_{j=1}^n (u_j + 1/u_j)\right)^2} \frac{du_1}{u_1} \cdots \frac{du_n}{u_n},$$

$$D_n := \frac{4}{n!} \int_0^\infty \cdots \int_0^\infty \frac{\prod_{i<j} \left(\frac{u_i - u_j}{u_i + u_j}\right)^2}{\left(\sum_{j=1}^n (u_j + 1/u_j)\right)^2} \frac{du_1}{u_1} \cdots \frac{du_n}{u_n}.$$

Consider the simplex $u_1 > u_2 > \cdots > u_n$. We can perform a change of variables via $u_k := \prod_{i=1}^k t_i$, with $t_1 \in (0, \infty)$ and other $t_i \in (0, 1)$. Define $w_k = \prod_{i=2}^k t_i$ and $v_k := \prod_{i=k}^n t_i$, and the functions

$$\mathcal{A}_n(t_2, t_3 \ldots, t_n) := \left(\prod_{n \geq k > j \geq 1} \frac{u_k/u_j - 1}{u_k/u_j + 1}\right)^2,$$

$$\mathcal{B}_n(t_2, t_3 \ldots, t_n) := \frac{1}{(1 + \sum_{k=2}^n w_k)(1 + \sum_{k=2}^n v_k)}.$$

Then the Ising expectation integrals can be recast as:

$$D_n = 2 \int_0^1 \cdots \int_0^1 \mathcal{A}\mathcal{B} \, dt_2 \, dt_3 \cdots dt_n, \tag{4}$$

$$C_n = 2 \int_0^1 \cdots \int_0^1 \mathcal{B} \, dt_2 \, dt_3 \cdots dt_n, \tag{5}$$

It is also useful to define

$$E_n := 2 \int_0^1 \cdots \int_0^1 \mathcal{A} \, dt_2 \, dt_3 \cdots dt_n. \tag{6}$$

As we noted earlier, at present there is no known practical scheme to find multi-hundred-digit values of general iterated integrals, except straightforward extensions of 1-D scheme that become impractically expensive beyond 3-D. Fortunately, we observed that in the case of the C_n integrals, we have

$$C_n = \frac{2^n}{n!} \int_0^\infty t K_0^n(t) \, dt, \tag{7}$$

where $K_0(t)$ denotes the *modified Bessel function* [1]. This converts the problem to a 1-D integral, thus dramatically reducing the computational cost.

We then computed 1000-digit values of C_n for selected values of n, ranging from 3 up to 1024 (corresponding to a 1024-fold iterated integral). With these numerical

values in hand, we were able to find (using PSLQ) and prove the following intriguing results [9]: $C_3 := L_{-3}(2)$ where $L_{-3}(s) = \sum_{n \geq 1}(1/(3n-2)^s - 1/(3n-1)^s)$, and $C_4 = 7\zeta(3)/12$, where $\zeta(s) := \sum_{k \geq 1} 1/k^s$ is the Riemann *zeta function* for $s > 1$. We further noticed that for large n, the numerical values appear to approach a constant, e.g.,

$$
\begin{aligned}
C_{10} &= 0.6318800241470122222903508736608083... \\
C_{40} &= 0.6304735033783635318699419018590694... \\
C_{100} &= 0.6304735033743867961220401927190317... \\
C_{200} &= 0.6304735033743867961220401927108789...
\end{aligned}
$$

By using the ISC tool, we immediately identified this constant:

$$
\lim_{n \to \infty} C_n = 2e^{-2\gamma},
$$

where γ is *Euler's constant*. We subsequently proved this intriguing numerical discovery [9], along with numerous extensions. Further research, experimental and theoretical, established the following results:

$$
\begin{aligned}
D_2 &= 1/3, \quad D_3 = 8 + 4\pi^2/3 - 27\,L_{-3}(2), \\
D_4 &= 4\pi^2/9 - 1/6 - 7\zeta(3)/2, \quad E_2 = 6 - 8\log 2, \\
E_3 &= 10 - 2\pi^2 - 8\log 2 + 32\log^2 2, \\
E_4 &= 22 - 82\zeta(3) - 24\log 2 + 176\log^2 2 - 256(\log^3 2)/3 \\
&\quad + 16\pi^2 \log 2 - 22\pi^2/3, \\
E_5 &\stackrel{?}{=} 42 - 1984\,\mathrm{Li}_4(1/2) + 189\pi^4/10 - 74\zeta(3) - 1272\zeta(3)\log 2 \\
&\quad + 40\pi^2 \log^2 2 - 62\pi^2/3 + 40(\pi^2 \log 2)/3 + 88\log^4 2 \\
&\quad + 464\log^2 2 - 40\log 2.
\end{aligned}
$$

In the case of E_5, the $\stackrel{?}{=}$ notation means that this is only an experimental discovery—as yet we do not have rigorous proof (and we may never have such a proof, because no proof may be accessible). Although in this case we were *not* able to reduce the integral to a 1-D form, we were nonetheless able to reduce its dimension by one by a certain transformation. The one-page-long 3-D integral that resulted from this manipulation is shown in [9]. The final evaluation of this integral to 250-digit accuracy required four hours on 64 CPUs of the Virginia Tech Apple system. Applying PSLQ to the resulting numerical value (together with the numerical values of a set of conjectured component terms), yielded the experimental evaluation shown above. Incidentally, the plot shown in Figure 1 is for the PSLQ run that discovered this identity.

The C_n constants also occur naturally within quantum field theory. Additional results on these integrals and some generalizations are given in [6] and [27].

4 Digit Randomness and Borel's Law

One of the oldest questions of mathematics is the whether constants such as π, e, $\sqrt{2}$ and $\log(2)$ have statistically random digits. Notably, von Neumann suggested the first computer-age calculation of π on the ENIAC for statistical reasons [19, Ch. 3]. More precisely, we can ask, for a given constant α, whether every m-long string of base-b digits appears, in the limit, with frequency b^{-m}. If so, then we say that α is *b-normal*. There is ample statistical evidence in favor of such conjectures—for instance, the constant π was recently computed to over one trillion digits (both in decimal and in base-16), but nothing was seen in statistical analysis of this data—see Table 1, which gives the simple digit distribution of π in decimal and in hexadecimal [19, Ch. 3].

Table 1. Digit distribution in base 10 and 16

Decimal Digit	Occurrences		Hex Digit	Occurrences
			0	62499881108
			1	62500212206
			2	62499924780
0	99999485134		3	62500188844
1	99999945664		4	62499807368
2	100000480057		5	62500007205
3	99999787805		6	62499925426
4	<u>100000</u>357857		7	62499878794
5	99999671008		8	<u>62500</u>216752
6	99999807503		9	62500120671
7	99999818723		A	62500266095
8	100000791469		B	62499955595
9	99999854780		C	62500188610
			D	62499613666
Total	**1,000,000,000,000**		E	62499875079
			F	62499937801

Total   **1,000,000,000,000**

   Along this line, it is reasonable to conjecture that every irrational root of an algebraic equation with integer coefficients is not only 10-normal, but is b-normal for all positive integers $b > 1$. Unfortunately, there are no proofs of normality for any of these well-known constants. We do not even know, provably, whether there are an infinite number of sevens in the decimal expansion of π, or whether the binary expansion of $\sqrt{2}$ has balanced frequencies of zeros and ones [19, Ch. 4]. Until recently, the only proofs of normality were for highly contrived constants such as *Champernowne's number*, namely $0.12345678910111213141516\ldots$ in base ten, or the corresponding concatenation in any given base [19, Ch. 4].

4.1 The BBP Formula for Pi

One interesting breakthrough in this area was the recent discovery of a formula that allows one to directly calculate binary digits of certain mathematical constants (including π and $\log 2$) beginning at an arbitrarily starting point, by means of a surprisingly simple algorithm. This algorithm is based on the following formula, which was discovered in 1996 by a computer program running the PSLQ integer relation algorithm extensively [11][4],[19, Section 3.4]:

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

It is easy to see how this individual digit-calculating scheme works by illustrating it for a similar formula, namely $\log 2 = \sum_{n=1}^{\infty} 1/(n2^n)$. Note that the binary expansion of $\log 2$ beginning after the first d binary digits is simply $\{2^d \log 2\}$, where $\{\cdot\}$ denotes fractional part. Thus we can write

$$\{2^d \log 2\} = \left\{ \sum_{n=1}^{\infty} \frac{2^{d-n}}{n} \right\} = \left\{ \sum_{n=1}^{d} \frac{2^{d-n}}{n} \right\} + \left\{ \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n} \right\}$$

$$= \left\{ \sum_{n=1}^{d} \frac{2^{d-n} \bmod n}{n} \right\} + \left\{ \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n} \right\}, \tag{9}$$

where we insert "mod n" in the numerator of the first term of (9) since we are only interested in the fractional part after division by n. Now the expression $2^{d-n} \bmod n$ may be evaluated very rapidly by means of the binary algorithm for exponentiation, where each multiplication is reduced modulo n. The entire scheme indicated by formula (9) can be implemented on a computer using ordinary 64-bit or 128-bit arithmetic—high-precision arithmetic software is not required. The resulting floating-point value, when expressed in binary format, gives the first few digits of the binary expansion of $\log 2$ beginning at position $d+1$. Similar calculations applied to each of the four terms in formula (8) yield a similar result for π. The largest computation of this type to date is binary digits of π beginning at the quadrillionth (10^{15}-th) binary digit, performed by an international network of computers organized by Colin Percival in the year 2000. This algorithm was also used to rapidly validate Kanada's 2002 computation of a trillion hexadecimal digits of π.

Since 1996, formulas of this type have been found for numerous other constants, using a similar experimental approach. A listing of many of these formulas is available in an online compendium [5]. No formula of this same type exists for π in any non-binary number base [21].

4.2 Statistical Implications of the BBP Formulas

One interesting (and unanticipated) discovery is that the existence of these computer-discovered BBP-type formulas has implications for the age-old question of normality

for several basic mathematical constants, including π and $\log 2$. This result, due to one of the present authors and Richard Crandall, is that if a mathematical constant has a BBP-type formula, then there is a related sequence which, if proven equidistributed in the unit interval, implies the normality of the constant. This means, for instance, that $\log 2$ is 2-normal if and only if the simple sequence defined by $x_0 = 0$ and $x_n = \{2x_{n-1} + 1/n\}$ is equidistributed in the unit interval. In a similar way, π is 2-normal if and only if the sequence $x_0 = 0$ and

$$x_n = \left\{ 16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21} \right\} \tag{10}$$

is equidistributed. Here the notation $\{\cdot\}$ means fractional part as before. Full details are given in [13] and [19, Section 3.8].

What's more, this line of research has recently led to a full-fledged proof of normality for an uncountably infinite class of explicit real numbers. Given $r \in [0, 1)$, let r_n be the n-th binary digit of r. Then for every r in the unit interval, the constant

$$\alpha_r = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n + r_n}} \tag{11}$$

is 2-normal and transcendental [14]. Further details on these results are given in [19, Sec. 4.3], [16], and [14].

Given the sequence (10) for π, we can define the hexadecimal (base-16) digit sequence $y_n = \lfloor 16x_n \rfloor$. In other words, we can divide the unit interval into 16 equal subintervals, labeled $(0, 1, 2, 3, \cdots, 15)$, and set y_n to be the label of the subinterval in which x_n lies. When this is done, a remarkable phenomenon occurs: The sequence (y_n) appears to perfectly (not just approximately) produce the hexadecimal expansion of π. In explicit computations, the first 1,000,000 hexadecimal digits generated by this sequence are identical with the first 1,000,000 hexadecimal digits of $\pi - 3$.

Evidently this phenomenon arises from the fact that in the sequence associated with π, the perturbation term $r_n = (120n^2 - 89n + 16)/(512n^4 - 1024n^3 + 712n^2 - 206n + 21)$ is summable, whereas the corresponding expression for $\log 2$, namely $1/n$, is not summable. In particular, define the sequence $\alpha_n = \{2^n\pi\}$, and let $||u - v||$ denote the distance between u and v in the wrapped unit interval. Then it can be shown that

$$||\alpha_n - x_n|| \leq \sum_{k=n+1}^{\infty} \frac{120k^2 - 89k + 16}{16^{j-n}(512k^4 - 1024k^3 + 712k^2 - 206k + 21)}$$
$$\approx \frac{120(n+1)^2 - 89(n+1) + 16}{16(512(n+1)^4 - 1024(n+1)^3 + 712(n+1)^2 - 206(n+1) + 21)},$$

so that $\sum_{n=1}^{\infty} ||\alpha_n - x_n|| \approx 0.01579..$

For the sake of heuristic argument, let us assume for the moment that the α_n are independent, uniformly distributed random variables in $(0, 1)$, and let $\delta_n =$

13

$||\alpha_n - x_n||$. Note that an error (i.e., an instance where x_n lies in a different subinterval of the unit interval than α_n) can only occur when α_n is within δ_n of one of the points $(0, 1/16, 2/16, \cdots, 15/16)$. Since $x_n < \alpha_n$ for all n (where $<$ is interpreted in the wrapped sense when x_n is slightly less than one), this event has probability $16\delta_n$. Then the fact that the sum 0.01579 has a finite value implies, by the first Borel-Cantelli lemma, that there can only be finitely many errors [18, pg. 153]. The comparable figure for $\log 2$ is infinite, which implies by the second Borel-Cantelli lemma that discrepancies can be expected to appear indefinitely, but with decreasing frequency.

Further, the small value of the sum suggests that it is unlikely that any errors will be observed. If instead of summing $||\alpha_n - x_n||$ from one to infinity, we instead sum from 1,000,001 to infinity (since we have computationally verified that there are no errors in the first 1,000,000 elements), then we obtain 1.465×10^{-8}, which suggests that it is very unlikely that any errors will ever occur. It would be nice to make "very unlikely" here more statistically precise.

5 Ramanujan Arithmetic-Geometric Continued Fractions

We now present some interesting results in the other direction—how data analysis and visualization methods typical of what is widely done in computational statistics can be employed in computational mathematics. Given $a, b, \eta > 0$, define

$$R_\eta(a, b) \;=\; \cfrac{a}{\eta + \cfrac{b^2}{\eta + \cfrac{4a^2}{\eta + \cfrac{9b^2}{\eta + \ddots}}}}.$$

This continued fraction arises in the *Notebooks* of the famous Indian mathematician Srinivasa Ramanujan (1887-1920). He discovered instances of the striking fact that

$$\frac{R_\eta(a, b) + R_\eta(b, a)}{2} \;=\; R_\eta\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

When the present authors first attempted to numerically compute $R_1(1, 1)$, only three reliable digits were obtained: $0.693\ldots$. But even at this accuracy, the result suggested that the value might be the well-known constant $\log 2$. In any event, higher accuracy was needed. From formula (1.11.70) of [20] one can see that for $0 < b < a$,

$$\mathcal{R}_1(a, b) \;=\; \frac{\pi}{2} \sum_{n \in \mathbb{Z}} \operatorname{sech} \frac{aK(k)}{K^2(k) + a^2 n^2 \pi^2} \operatorname{sech}\left(n\pi \frac{K(k')}{K(k)}\right), \qquad (12)$$

14

where $k = b/a = \theta_2^2/\theta_3^2, k' = \sqrt{1-k^2}$. Here θ_2, θ_3 are *Jacobian theta functions* [1, Section 16.27], and K is a complete elliptic integral of the first kind [1, Section 17.3].

Viewing the previous equation as the limit of a Riemann sum, we have

$$\mathcal{R}(a) := \mathcal{R}_1(a,a) \quad = \quad \int_0^\infty \frac{\text{sech}(\pi x/(2a))}{1+x^2} dx \quad = \quad 2a \sum_{k=1}^\infty \frac{(-1)^{k+1}}{1+(2k-1)a}, \quad (13)$$

where the final equality follows from the Cauchy-Lindelöf Theorem. This sum may also be written in terms of hypergeometric functions [1, Section 15] as $\mathcal{R}(a) = \frac{2a}{1+a} F\left(\frac{1}{2a} + \frac{1}{2}, 1; \frac{1}{2a} + \frac{3}{2}; -1\right)$. The latter form can be used in *Maple* or *Mathematica* to determine that

$$\mathcal{R}(2) \quad = \quad 0.974990988798722096719900334529\ldots.$$

The ISC tool (or 'identify' in *Maple*) reports this constant is equal to $\psi(3/8) - \psi(7/8)$, where $\psi(t) := \Gamma'(t)/\Gamma(t)$ is the *digamma* function [1, Section 6.3]. With some additional manipulation, one obtains the simpler form $\mathcal{R}(2) = \sqrt{2}[\pi/2 - \log(1+\sqrt{2})]$.

Indeed, it can now be seen that [24]:

$$\mathcal{R}(a) \quad = \quad 2 \int_0^1 \frac{t^{1/a}}{1+t^2}\, dt.$$

Note that $\mathcal{R}(1) = \log 2$. No non-trivial closed form is known for $\mathcal{R}(a,b)$ with $a \neq b$, although

$$\mathcal{R}_1\left(\frac{1}{4\pi} \beta\left(\frac{1}{4}, \frac{1}{4}\right), \frac{\sqrt{2}}{8\pi} \beta\left(\frac{1}{4}, \frac{1}{4}\right)\right) \quad = \quad \frac{1}{2} \sum_{n \in \mathbf{Z}} \frac{\text{sech}(n\pi)}{1+n^2},$$

comes close. Here β denotes the classical *Beta function* [1, Section 6.2]. Further details are to be found in [24, 23, 20].

Study of these Ramanujan continued fractions was further facilitated by examining the closely related dynamical system $t_0 = 1, t_1 = 1$ and

$$t_n := t_n(a,b) \quad = \quad \frac{1}{n} + \omega_{n-1}\left(1 - \frac{1}{n}\right) t_{n-2} \quad (14)$$

where $\omega_n = a^2$ or b^2 (from the Ramanujan continued fraction definition), depending on whether n is even or odd. If we inspect this recursion based only on numerical values, nothing much is evident—one only notes that $t_n \to 0$ fairly slowly. If, however, we look at this iteration pictorially, we learn significantly more—see Figure 2.

If we then scale by $\sqrt{n}$ (as was suggested by theoretical considerations but is easily suggested numerically) and color the iterations blue or red depending on whether n is odd or even, then some remarkable fine structure appears—see Figure 3. With assistance of such plots, the behavior of these iterates (and the Ramanujan

15
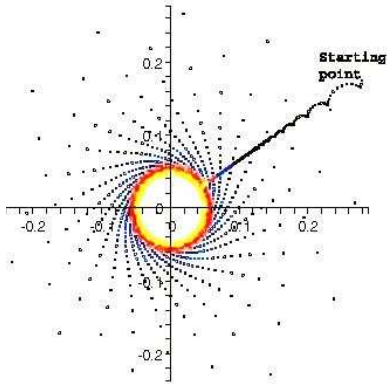
In particular, if we plot (14) in the complex plane changing color every few hundred iterates we see dramatic evidence of structure.

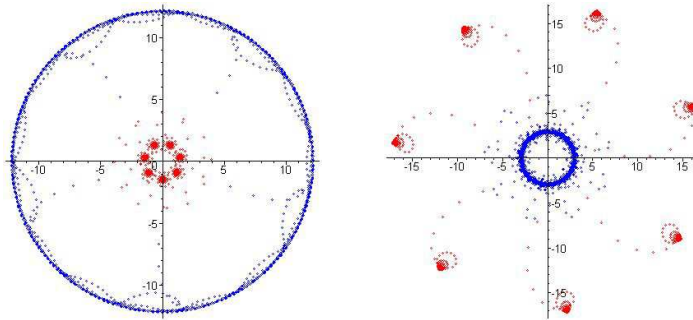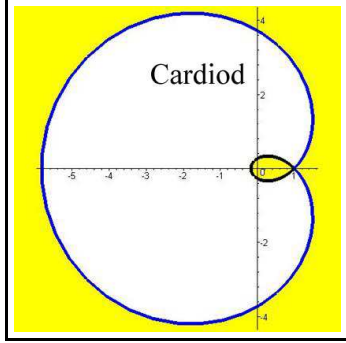Figure 2: The first few thousand iterates of 14 plotted in the complex plane

Figure 3: Dynamics and attractors of two different cases

A scatter plot of several thousands of points for which numerical convergence seemed to occur. The picture was accurate enough to identify the exact equation of the cardioid. Remarkably the cardioid is precisely the domain of validity of the complex AG inequality.

Figure 4: A cardioid revealed

continued fractions) are now quite well understood. These studies have ventured into matrix theory, real analysis and even the study of martingales from probability theory [24, 23, 22, 26]. The exact domain of convergence in the complex plane of the original continued fraction is the cardioid $C := \{c : |1 + c|)/2 \geq \sqrt{|c|}\}$, where $c := b/a$—see Figure 4.

There are exceptional cases when a/b is purely imaginary. *Jacobsen-Masson theory* [23, 22] shows that the even/odd fractions for $\mathcal{R}_1(i, i)$ behave "chaotically" and do not converge. Indeed, when $a = b = i$, the values $(t_n(i, i))$ exhibit a fourfold quasi-oscillation, as n runs through values mod 4. Plotted versus n, the (real) sequence $t_n(\mathbf{i})$ exhibits the serpentine oscillation of four separate "necklaces," as shown in Figure 5. The detailed asymptotic is

$$t_n(i, i) = \sqrt{\frac{2}{\pi} \cosh \frac{\pi}{2}} \frac{1}{\sqrt{n}} \left(1 + O\left(\frac{1}{n}\right)\right) \times \begin{cases} (-1)^{n/2} \cos(\theta - \log(2n)/2) \\ (-1)^{(n+1)/2} \sin(\theta - \log(2n)/2), \end{cases}$$

where the first form is taken when n is even, and the second when n is odd. Here $\theta = \arg \Gamma((1 + i)/2)$.

Analysis is remarkably easy once given the following striking hypergeometric parameterization of (14) when $a = b \neq 0$, see [22], which was both discovered and proved largely by computer:

$$t_n(a, a) = \frac{1}{2} F_n(a) + \frac{1}{2} F_n(-a), \tag{15}$$

where $F_n(a)$ is defined by:

$$F_n(a) := -\frac{a^n 2^{1-\omega}}{\omega \, \beta(n + \omega, -\omega)} {}_2\mathrm{F}_1\left(\omega, \omega; n + 1 + \omega; \frac{1}{2}\right).$$

Here ${}_2\mathrm{F}_1$ denotes the Gauss hypergeometric function [1], and

$$\beta(n + 1 + \omega, -\omega) = \frac{\Gamma(n + 1)}{\Gamma(n + 1 + \omega) \Gamma(-\omega)},$$
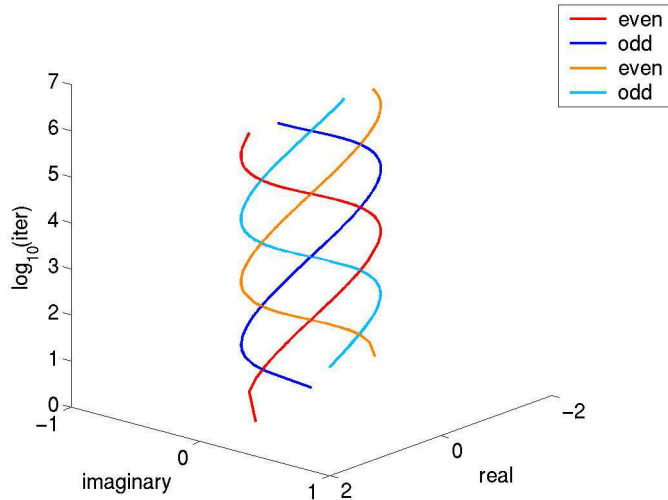
17

Figure 5: The subtle four fold serpent

is a Beta-function value, where $\omega := (1 - 1/a)/2$. Indeed, once (15) was discovered by a combination of insight and methodical computer experiment, its proof is highly representative of the changing paradigm: both sides satisfy the same recursion and the same initial conditions. This can be checked in *Maple* and if one looks inside the computation, one learns which *confluent hypergeometric identities* are needed for an explicit human proof.

6 Other topics

We finish by mentioning a number of other computational topics which the reader can follow up on in the given references.

(a) *Entropic measures of algorithmic complexity.* In [25] an analysis is performed of the binary expansion of $\Gamma(1/n), n = 2, 3, \ldots, 6$. For all cases except $n = 5$ fast algorithms are known [19, Ch. 6]. For $n = 5$ it is conjectured no such algorithm exists and the computation of various entropy measures supports this dichotomy.

(b) *Data mining.* Resources like the *Online encyclopedia of integer sequences*

$$\texttt{http:///www.research.att.com/\textasciitilde njas/sequences} \qquad .$$

allow for substantial data mining [19, Ch. 8] [3]. Tools to automate the process are still quite primitive. A variety of such resources is documented at

$$\texttt{http://www.experimentalmath.info} \qquad .$$

(c) *Automated proof techniques.* There are well-developed algorithms for automating the proof of identities of hypergeometric type due to Gosper and to Wilf-Zeilberger among others [3, 7]. Equally potent tools for discovery of gen-

18

erating functions, recursions and differential equations are described in [27] and [19, 20, 3]. Less well-known are algorithms for automated proof of many inequalities (see the article by Kauers in the same volume as [7]).

7   Conclusion

In the above survey, while we have focused on our own work and that of our collaborators, we are of course not alone in adopting these techniques. Many examples originating with other researchers are scattered throughout [ema,expm1,expm2]. In particular [ema, p. 190] makes a concise review of early work on and writings about computer experiments in mathematics. Another good reference is the recent book by Villegas, who employs a similar experimental approach [29].

We have seen how the emerging discipline of experimental mathematics has much in common with computational statistics. Both disciplines seek to extract underlying analytic facts out of "raw" data. Both disciplines involve ever-more sophisticated mathematical and computational techniques. Both disciplines have applications to each other and to numerous other fields of science. This article is written in the spirit of fostering greater interaction between these two communities.

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1970.

[2] Chris Anderson. The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 16(7), 2008.

[3] D. H. Bailey, J. M. Borwein, N. J. Calkin, R. Girgensohn, D. R. Luke, and V.H. Moll. *Experimental Mathematics in Action*. A K Peters, Wellesley, MA, 2004.

[4] David H. Bailey. The BBP formula for pi (2006). `http://crd.lbl.gov/~dhbailey/dhbpapers/bbp-alg.pdf`.

[5] David H. Bailey. A compendium of BBP-type formulas (2000). `http://crd.lbl.gov/~dhbailey/dhbpapers/bbp-formulas.pdf`.

[6] David H. Bailey, David Borwein, Jonathan M. Borwein, and Richard Crandall. Hypergeometric forms for ising-class integrals. *Experimental Mathematics*, 16(3):257–276, 2007.

[7] David H. Bailey and Jonathan M. Borwein. Computer-assisted discovery and proof. In Victor Moll, editor, *Tapas in Experimental Mathematics*, Contemporary Mathematics, pages 21–52. American Mathematical Society, 2008.

[8] David H. Bailey and Jonathan M. Borwein. Highly parallel, high-precision numerical integration. *International Journal of Computational Science and Engineering*, to appear, 2008.

[9] David H. Bailey, Jonathan M. Borwein, and Richard E. Crandall. Integrals of the ising class. *Journal of Physics A: Mathematical and General*, 39:12271–12302, 2006.

[10] David H. Bailey, Jonathan M. Borwein, and Richard E. Crandall. Box integrals. *Journal of Computational and Applied Mathematics*, 206:196–208, 2007.

[11] David H. Bailey, Peter B. Borwein, and Simon Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation*, 66:903–913, 1997.

[12] David H. Bailey and David J. Broadhurst. Parallel integer relation detection: Techniques and applications. *Mathematics of Computation*, 70:1719–1736, 2000.

[13] David H. Bailey and Richard E. Crandall. On the random character of fundamental constant expansions. *Experimental Mathematics*, 10:175–190, 2001.

[14] David H. Bailey and Richard E. Crandall. Random generators and normal numbers. *Experimental Mathematics*, 11:527–546, 2003.

[15] David H. Bailey, Xiaoye S. Li, and Karthik Jeyabalan. A comparison of three high-precision quadrature schemes. *Experimental Mathematics*, 14(3):317–329, 2005.

[16] David H. Bailey and Michal Misiurewicz. A strong hot spot theorem. *Proceedings of the American Mathematical Society*, 134:2495–2501, 2006.

[17] Michael S. Becker. Problem 11275. *American Mathematical Monthly*, 11(2):165, 2007.

[18] Patrick Billingsley. *Probability and Measure*. John Wiley, Hoboken, NJ, 1986.

[19] Jonathan M. Borwein and David H. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. A K Peters, Wellesley, MA, 2004, expanded second edition, 2008.

[20] Jonathan M. Borwein, David H Bailey, and Roland Girgensohn. *Experimentation in Mathematics: Computational Paths to Discovery*. A K Peters, Wellesley, MA, 2004.

[21] Jonathan M. Borwein, David Borwein, and William F. Galway. Finding and excluding b-ary Machin-type BBP formulae. *Canadian Journal of Mathematics*, 56:897–925, 2004.

[22] Jonathan M. Borwein, Richard Crandall, David Borwein, and Raymond Mayer. On the dynamics of certain recurrence relations. *Ramanujan Journal*, 13:63–101, 2007.

[23] Jonathan M. Borwein and Richard E. Crandall. On the Ramanujan AGM fraction. Part II: the complex-parameter case. *Experimental Mathematics*, 13(3):287–296, 2004.

[24] Jonathan M. Borwein, Richard E. Crandall, and Greg Fee. On the Ramanujan AGM fraction. Part I: the real-parameter case. *Experimental Mathematics*, 13(3):275–286, 2004.

[25] Jonathan M. Borwein and Kostas Karamanos. Algebraic dynamics of certain Gamma function values. In *Generalized Convexity and Generalized Monotonicity*, volume 77 of *Nonconvex Optimization and its Applications*. Springer-Verlag, New York, 3–21.

[26] Jonathan M. Borwein and D. Russell Luke. Dynamics of some random continued fractions. *Abstract and Applied Analysis*, 5:449–468, 2005.

[27] Jonathan M. Borwein and Bruno Salvy. A proof of a recursion for Bessel moments. *Experimental Mathematics*, 17:223–230, 2008.

[28] Kevin Kelly. The end of theory. *Edge*, (249), July 2008.

[29] Fernando Villegas. *Experimental Number Theory*. Oxford University Press, Oxford, UK, 2007.