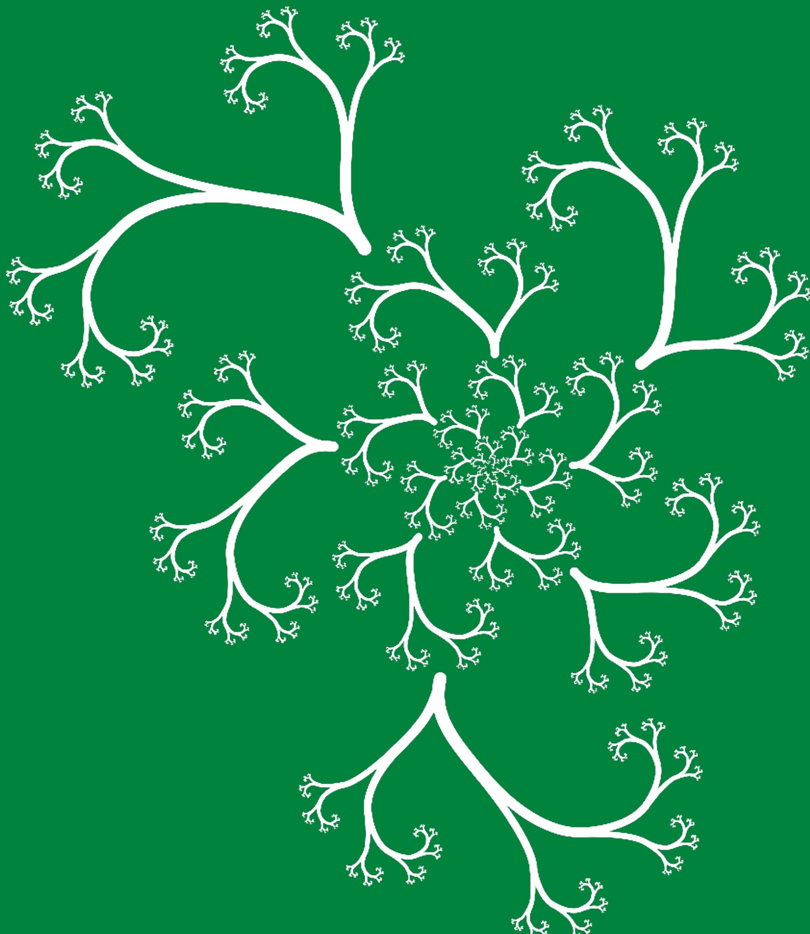
Vol. 90, No. 3, June 2017

MATHEMATICS MAGAZINE



- An optimal group testing algorithm and Fibonacci numbers
- Differentiating first and second serves in tennis
- Memoryless processes, random walks, and hypergeometric functions, all from a game
- Solving a new class of inexact differential equations with Fontaine's method

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. The *Magazine* is not a research journal, so the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for the Magazine. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships among various branches of mathematics and between mathematics and other disciplines.

Submissions of articles are required via the *Mathematics Magazine*'s Editorial Manager System. The name(s) of the author(s) should not appear in the file. Initial submissions in pdf or LaTeX form can be sent to the editor at www.editorialmanager.com/mathmag/.

The Editorial Manager System will cue the author for all required information concerning the paper. Questions concerning submission of papers can be addressed to the editor at mathmag@maa.org. Authors who use LaTeX are urged to use the Magazine article template. However, a LaTeX file that uses a generic article class with no custom formatting is acceptable. The template and the Guidelines for Authors can be downloaded from www.maa.org/pubs/mathmag.

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, NW, Washington, DC 20036 and Lancaster, PA, in the months of February, April, June, October, and December.

Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Address advertising correspondence to

MAA Advertising
1529 Eighteenth St. NW
Washington, DC 20036
Phone: (202) 319-8461
E-mail: advertising@maa.org

Further advertising information can be found online at www.maa.org.

Change of address, missing issue inquiries, and other subscription correspondence can be sent to:

The MAA Customer Service Center
P.O. Box 91112
Washington, DC 20090-1112
(800) 331-1622
(301) 617-7800
maaservice@maa.org

Periodicals postage paid at Washington, D.C. and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, NW, Washington, DC 20036-1385.

Printed in the United States of America.

COVER IMAGE

Phyllotactic Filigree © 2017 David A. Reimann (*Albion College*). Used by permission.

This work extensively uses the golden ratio, $\phi = \frac{1}{2}(1 + \sqrt{5})$, for rotations and scalings between components.

Vol. 90, No. 3, June 2017

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

The first article in this issue is by Seth Zimmerman. Zimmerman explains group testing, the practice of grouping samples to minimize the number of tests necessary to determine who has a disease in a large population, constructs a group testing algorithm, and proves its optimality, providing what is believed to be the first proof of optimality of such an algorithm. He also notices a relationship to the Fibonacci numbers.

The Pythagorean theorem is perhaps the most often proved result in mathematics, or at least most submitted to THIS MAGAZINE. Kevin Ferland extends two classic proofs of the Pythagorean theorem to the law of cosines, while Patrik Nystedt uses the law of cosines to prove the law of sines. Nam Gu Heo uses Ptolemy's theorem in a proof without words to prove the Pythagorean theorem.

For sports fans, May/June has the French Open while July has Wimbledon. Yigal Gerchak and Marc Kilgour consider the optimal serving strategy for first and second serves in tennis. Their paper takes as its starting point a 1971 article by David Gale that appeared in THIS MAGAZINE. I'll report back in the fall if my tennis performance in a summer mixed doubles league improves.

I always enjoy when a problem can be attacked using different types of mathematics, like a many faceted diamond can be viewed from different angles. Steven Miller and a cast of students introduce a simple game based on a simple probabilistic rule. Their article analyzes the game and highlights connections to memoryless processes, random walks on lattices, statistical inference, and hypergeometric functions.

The modern integrating factor method to solve ordinary differential equations dates back to a paper by Clairaut from 1739. He was motivated in part by a paper by Fontaine. Aaron E. Leanhardt and Adam E. Parker revisit Fontaine's method and its history, while providing what they believe is the first family of ordinary differential equations that is solvable by Fontaine's method but has no obvious solution using the modern technique.

I find it interesting to think about a mathematical result from a particular point of view. Exploiting this idea, K. K. Kataria looks at a combinatorial result from a probabilistic perspective by using probability to prove the multinomial theorem. As a corollary of an initial result, he shows that the number of nonnegative integer solutions to $x_1 + \cdots + x_m = k$ is a binomial coefficient–a favorite of mine.

Jeffrey Bergen provides a one-page proof that nth roots are always integers or irrational. His proof depends on the well ordering principle and the division algorithm. He asks whether this is the easiest proof of this result. Claudi Alsina and Roger Nelsen use rectangular boxes to visualize a proof of Maclaurin's inequality. Mixed in between the articles are proofs without words by Victor Oxman and Moshe Stupel and by Ángel Plaza.

There are two puzzles in this issue: Brendan Sullivan offers a crossword about MathFest while Lai Van Duc Thinh provides another Pinemi puzzle. Note that due to a production error, there was a mistake in the April Pinemi puzzle. There is a boxed entry on page 181 explaining the error; our apologies. Also, be on the lookout for boxed entries throughout the issue about the mathematics on Snapple caps. The issue concludes with the Problems and Reviews.

Michael A. Jones, Editor

ARTICLES

Detecting Deficiencies: An Optimal Group Testing Algorithm

SETH ZIMMERMAN
Bellingham WA 98229
ziml s@earthlink.net

Group testing has been the subject of continuing investigation since 1943, when the U.S. Army was looking for an economical way to screen recruits for syphilis by pooling minute blood samples [**1**]. Consider this scenario, commonly used for comparing results: We are testing samples of blood from one million people, attempting to locate the carriers of a disease known to be present in 0.01% of the population. With no further information, we assume each person has the same probability of testing positive (being diseased). Instead of performing one million individual tests, we might economize by pooling samples. For example, if we test a mixture of 100 samples, the result is likely to be negative, and 99 tests will have been avoided. On the other hand, testing a mixture of all one million blood samples would be pointless since several of the included samples would, no doubt, be positive, and the test would yield no new information. If a requirement is to find all positive samples, can one specify an optimum procedure to do so, that is, one that, for any set of samples, finds all positives with the minimum expected number of tests? Such a procedure requires a set of operating rules that specify not only how many samples should be grouped together in the initial tests but also how to proceed within a group if it tests positive.

Since its introduction in the above form, mathematically equivalent situations having to do with production lines, computer wiring, DNA screening, and other areas have produced a variety of approaches [**2**, pp. 1–5, **6, 8**]. While blood testing is perhaps the version most easily grasped, we can state our assumptions more generically:

(i) Tests are being administered to a large population of n samples in order to determine exactly which samples test positive.

(ii) Each sample has the same probability p of testing positive.

(iii) A test of a group of samples will register positive whenever one or more of the samples included in it are positive.

(iv) The cost of testing a group is always one unit, regardless of group size.

(v) If by previous tests a group is known to contain a positive sample, the group need not be tested, although subgroups within it might be.

(vi) Once a group is tested, a subsequent test that includes samples within that group cannot include samples outside that group. This is typically referred to as "nesting."

Within the constraints of these six assumptions, there are several ways to define a valid search procedure. In this paper, a valid search procedure is the following.

Math. Mag. **90** (2017) 167–178. doi:10.4169/math.mag.90.3.167. © Mathematical Association of America
MSC: Primary 62K05, Secondary 68W40; 68R05; 11B39

Definition 1. A valid search procedure is defined as one that proceeds according to a predetermined sequence of tests, the only exception being the omission of any test known to be positive as a result of earlier tests in the sequence.

Some Points to be Noted

Considering all the attention given to this now classic problem, one might well ask why a further approach is needed. Indeed, four high school calculus students produced four reasonably efficient methods for treating an example just slightly different from the one specified above [**9**]. The attraction of course lies not in trying to produce a new method to lower the current minimum number of tests in this example. Indeed, for any economic or industrial application, we are probably close enough to optimum. The interest here is purely mathematical—the challenge of either producing an optimum procedure or proving that one does not exist.

(Spoiler alert: Before we reveal anything in the next paragraph, the reader might like to write down his/her idea for an optimum method—at least how large the initial groups to be tested should be.)

In contrast to previous top-down methods, this paper takes a constructive, algebraic approach whose initial investigation can be found in [**11**]. While the algorithm in this paper appears to be the first proven optimum procedure, it is restricted to the above six assumptions and our version of the valid search procedure. Although this is the most common formulation of the problem, there are others that we will discuss below. After presenting the algorithm (whose proof is detailed in the last section), we will explore the Fibonacci pattern that consistently emerges.

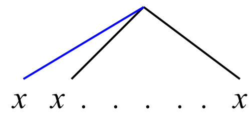
The traditional terminology for this problem uses "group" to signify a set of samples, rather than its standard algebraic definition.

Most papers in the literature consider a version of the problem in which a specific number of positive samples is already known to exist, rather than assuming the same probability for each sample. This is sometimes realistic, but more often not. In this paper, we know the probability of any individual sample's registering positive, an assumption that carries less information than knowledge of the specific number of positives.

Terminology and Method in Simple Cases

Two samples Consider the simplest nontrivial case, that of two samples. There are two possible procedures for testing. The first is to test each sample separately, which we represent by xx. The second is to test them together and then, if a defect is registered, test them separately. A common way of representing this would be a rooted tree graph with two branches. While the testing procedure for any number of samples can always be represented by a tree graph, this becomes cumbersome as that number increases. For this reason, we will denote the second procedure— testing two samples together and then, if a defect is registered, testing them separately—by $\overline{xx}$, where the horizontal line signifies a test of the two samples together. When extended to situations with several samples, this representation will make the testing order readily decipherable. The horizontal lines can be seen as space-saving versions of the graph edges. To be precise we give the following definition.

Definition 2. A test of several samples together, followed by individual tests of each sample, conventionally represented by the tree graph, will be represented by $\overline{xx\ldots x}$. Similar representation will be used for all testing situations.



Since the expected number of tests for xx is 2, regardless of p, we wish to know when $\overline{xx}$ is lower than 2. The four possibilities for a pair of samples are: [positive, positive], [positive, negative], [negative, positive], and [negative, negative]. For convenience, instead of the words "positive" and "negative," we can use the probabilities p and q themselves, where $q = 1 - p$. We will also place to the right of each possibility not the number of tests entailed by that possibility, but rather its excess above or advantage below two tests. Thus, for $\overline{xx}$:

$$[p, p]1 \quad [p, q]1 \quad [q, p]0 \quad [q, q]{-1}$$

For example, $[p, q]$ will require three tests—both together, then the first sample, then the second—so its excess is 1. $[q, p]$ will require only two tests—both together, then the first sample, revealing that the second sample is p.

We now compute $V[\overline{xx}]$, the "value" of the procedure $\overline{xx}$, that is, the expected number of tests minus 2. (The general definition of value is given in the next section.) Multiplying the probability of each of the four possibilities above by its excess or advantage, we compute

$$V[\overline{xx}] = p^2 + pq + 0 - q^2 = (1-q)^2 + q(1-q) - q^2 = 1 - q - q^2.$$

If the value is negative, then $\overline{xx}$ is the better choice. If the value is positive, then xx is better. Setting the above polynomial in q equal to zero and confining ourselves to the interval [0,1], we find that $\frac{\sqrt{5}-1}{2}$ or φ, the reciprocal of the golden mean, is the dividing point. If $q > \varphi$ (≈ 0.618), then $\overline{xx}$ is better than xx, that is, has a lower value.

Since group testing is applicable only if p is quite small (q near 1), we can assume from now on that $q > \varphi$. In addition, it will be convenient to write all of our equations, inequalities, and expressions in terms of q.

Three samples Testing three samples individually is represented by xxx and, by analogy with the case of two, is given the value 0. Testing two together and one separately is represented by either $x\overline{xx}$ or $\overline{xx}x$, and from the results of the next section, we see that these have the same value, $V[x\overline{xx}] = V[\overline{xx}x] = 1 - q - q^2$. (Here, the value is the expected number of tests minus 3 since testing separately would require three tests.)

One might perform $\overline{xxx}$, that is, test all three together and then, if necessary, test individually. Omitting the computation, $V[\overline{xxx}] = 1 - q^2 - 2q^3$. A comparison of $V[\overline{xxx}]$ and $V[\overline{x\overline{xx}}]$ shows that $\overline{xxx}$ can never be optimal.

n samples

Definition 3. For any number of samples, n, a structure is defined as an arrangement of x's and horizontal lines, where each line represents a collective test of all the samples beneath it. Tests are conducted from top to bottom and left to right.

A structure unambiguously describes the procedure for testing n samples. For instance, an inclusive test of four samples, followed by a test of the first alone, then the last three, then the second alone, then the last two, then the third and the fourth—each test performed only when its outcome is not already known from previous tests—would be represented by $\overline{x\overline{x\overline{xx}}}$. (There will be no ambiguity in context if we use the word "test" to mean either an actual test or the line signifying a test.)

Definition 4. The value of a structure, $V[]$, is defined as its expected number of tests minus n, the number of samples it includes.

Our goal in the next three sections, stated once more, is to produce an algorithm that, for any given n and q, generates the structure with the lowest expected number of tests. As with the cases of $n = 2$ and $n = 3$, this is equivalent to finding the structure of least value.

Definition 5. Test A is said to include, or is inclusive of, test B if all samples included in test B are included in test A. An all-inclusive test on a structure is defined as a test that includes all of its samples and is represented by a line above all other lines in the structure.

What change in value occurs when an all-inclusive test is added to a given structure that lacks one? That is, what is V[the structure with the all-inclusive test] $-V$[the structure without the all-inclusive test]? Here we assume that, if the all-inclusive test registers positive, we continue testing exactly as we would have tested without it.

Consider structure #1, represented by $\overline{n_1}\,\overline{n_2}\ldots\overline{n_k}$, where $\sum_{i=1}^{k} n_i = n$. In this structure, there are k groups of samples, and we are representing each group simply by the number of samples in that group, even though each group may have its own substructure. We need not specify these substructures for we will soon see that the value of adding an all-inclusive test above all k groups depends only upon the number of groups, k, and the number of samples, n_i , in each group, not upon their precise substructures. Thus, $\overline{n_i}$ represents a test of n_i x's with an unspecified substructure. If we add an all-inclusive test to structure #1, we will produce structure #2: $\overline{\overline{n_1}\,\overline{n_2}\ldots\overline{n_k}}$.

Since we have added a single all-inclusive test to structure #1 to obtain structure #2, the value appears to increase by 1. However, should all the samples in the first (k–1) groups be negative and some sample in group k be positive, there would be no need to test group k since, after testing the (k-1) groups that are negative, group k would have to be positive. In this case, the value would not increase by one but remain the same. Thus, from an increase of 1, we should subtract the probability that this might occur, namely $\left(q^{\sum_{i=1}^{k-1} n_i}\right)(1 - q^{n_k})$ or $(q^{n-n_k})(1 - q^{n_k})$. In addition, should all n samples be negative, we would not have to perform any of the tests included in the k groups. This would occur with probability q^n, and therefore, we must subtract a further kq^n. The total additional value as a result of adding an all-inclusive test is then

$$1 - (q^{n-n_k})(1 - q^{n_k}) - kq^n = 1 - q^{n-n_k} - (k - 1)q^n.$$

For convenience, we refer to this last expression, $1 - q^{n-n_k} - (k - 1)q^n$, as E_1. What we have just shown is the following theorem.

Theorem 1. *The value of adding an all-inclusive test above a structure consisting of k groups depends only upon the number of groups, k, and the number of samples, n_i, in each group; it is independent of the precise substructures of these k groups.*

We now have a rapid method for computing the value of any structure. Instead of going through the 2^n possibilities, as we did for $n = 2$ and 3, we simply begin with n individual samples (value equal zero) and consider the tests one by one as they grow more inclusive, that is, as they contain more and more tests beneath them. With each test, we add the amount given by E_1, an amount we can call the value of the test. When no more tests remain, we have calculated the total value of the structure.

Note that the order in which we calculate the value of a structure is the reverse of the order in which we execute the physical tests represented by that structure.

Two Theorems Concerning Optimal Structures

Definition 6. If a test has a negative value, it is said to be advantageous; if it has a positive value, it is said to be disadvantageous.

The two theorems in this section will prove that, in building a structure on n samples from the least inclusive tests up through the more inclusive—with the intention of finding the structure of least value—a disadvantageous test should never be added, and a test should never be applied to more than two tests immediately below it. That is, the optimal structure will contain only advantageous tests, and the structure will be binary, equivalent to a binary rooted tree. The proofs are given in the "Proofs" section below.

Theorem 2. *In building any structure, optimal or not, if a test is immediately disadvantageous, it can never be eventually advantageous.*

This theorem states that, if a test has a positive value, it should not be added. It is tempting to speculate that a disadvantageous test might eventually prove advantageous once inclusive tests are added above it, but this is never the case.

Theorem 3. *Within an optimal structure, any test on m samples ($m > 1$) must include precisely two tests directly beneath it, where these two tests together include those same m samples. Thus, the structure $\overline{\overline{m_1}\,\overline{m_2}\ldots\overline{m_k}}$ could appear in an optimal structure only if $k = 2$.*

It should be noted that only now, using Theorems 2 and 3, is it possible to formally prove that for any n, if $q \leqslant \varphi$, the structure without any group test at all is optimal.

The Optimal Search Algorithm

We can now specify a recursive algorithm to obtain the optimal structure on n samples. The structure thus obtained determines, in nesting fashion, the physical tests to be performed in the actual search. The proof that it is optimal is presented in the final section.

Definition 7.

a) $O_q(r)$ is defined as the optimal structure on r samples, each with the same probability q of being negative; $V_q(r)$ is its value.

b) The union of two structures, $S_1 \bigcup S_2$, is defined as the structure obtained by placing S_2 to the right of S_1 and considering this a new structure.

c) $\overline{S_1 \bigcup S_2}$ is defined as $S_1 \bigcup S_2$ with the addition of an inclusive test of all samples.

d) A_1 is defined as the set of structures defined by

$$\left\{ O_q(n') \bigcup O_q(n - n') : 1 \leqslant n' \leqslant \left\lfloor \frac{n}{2} \right\rfloor \right\}.$$

e) A_2 is defined as the set of structures defined by

$$\left\{ \overline{O_q(n') \bigcup O_q(n - n')} : 1 \leqslant n' \leqslant \left\lfloor \frac{n}{2} \right\rfloor \right\}.$$

Theorem 4. *$O_q(r)$ is the structure of least value $V_q(r)$, selected from all structures included in A_1 or A_2.*

Note that the set A_2 consists of the same $\lceil n/2 \rceil$ structures as those in A_1, with an additional all-inclusive test. Thus, utilizing the expression E_1 above, if the value of the n'th structure in A_1 is $V_q(n') + V_q(n - n')$, the value of the n'th structure inA_2 is $V_q(n') + V_q(n - n') + 1 - q^{n'} - q^n$.

To briefly illustrate, imagine that in our recursive search for $O_q(7)$ we already know the optimal structures $O_q(n)$, and their values $V_q(n)$ for n $\leqslant 6$. We now choose the minimum of the values of the six structures $O_q(1) \bigcup O_q(6)$; $O_q(2) \bigcup O_q(5)$; $O_q(3) \bigcup O_q(4)$; $\overline{O_q(1) \bigcup O_q(6)}$; $\overline{O_q(2) \bigcup O_q(5)}$; $\overline{O_q(3) \bigcup O_q(4)}$. This is $V_q(7)$, and the structure that produced it is $O_q(7)$.

In summary, we begin by obtaining $O_q(2)$ and its value $V_q(2)$. From this, we obtain $O_q(3)$ and its value $V_q(3)$ and so on recursively until we reach $O_q(n)$ and its value $V_q(n)$. A structure of tests is thus recursively built, always combining exactly two tests, until a further test is no longer advantageous. The reader might like to get a sense of this by using a hand calculator to obtain $O_q(6)$ and $V_q(6)$ for q = 0.9999, which can be compared with Table 1 below. To begin, $O_q(2)$ and $V_q(2)$ are readily calculated using $1 - q - q^2$. $O_q(3)$ then has only one possibility, and $V_q(3)$ can be obtained using $1 - q - q^3$. $O_q(4)$ has two potential divisions into 1, 3 or 2, 2, and for both of these, four separate calculations must be made—with and without an inclusive test above—and the lowest of these becomes $V_q(4)$. Two more such procedures finally yield $O_q(6)$ and $V_q(6)$.

In practice, computing time can be conserved by noting that, since a test can never include more than two tested groups immediately beneath it, its value over n samples will be of the form $1 - q^m - q^n$. This has a minimum of $1 - q - q^n$, when $m = 1$, so we never need consider a test of more than $n_{\max}$ samples, where $n_{\max}$ is the greatest n such that $1 - q - q^n \leqslant 0$. Solving for u in $1 - q - q^u = 0$, it follows that $u = \log(1 - q)/\log(q)$ so that $n_{\max} = \lceil \log(1 - q)/\log(q) \rceil$. While this is convenient as a time saver, a more exact definition and calculation of $n_{\max}$, known as the "cutoff point," has been known for many years [**4, 10**].

Computational Results and the Conjectured Fibonacci Pattern

A typical example In an optimal structure, if the size of the population, n, is sufficiently great, the number of samples included in the most inclusive test is a function of q only; n is irrelevant. For example, in an optimal structure for $q = 0.9999$, the most inclusive tests are in groups of 6765. For a population with $n > 6765$, a test that included more than 6765 samples would always be disadvantageous. To continue this particular example, if any such group of 6765 samples registers positive, the subgroups to be tested should have sizes 2584 and 4181. Successive divisions, if positives are registered, are listed in Table 1, together with the expected number of tests. (The highlighted entries suggest the pattern conjectured below.) With $q = 0.9999$ and $n = 1{,}000{,}000$, the program took several minutes to run using Maple 10 on a Macintosh running OS X.

As stated above, the table assumes $q = 0.9999$ for various group sizes, n, whether that group is an entire population or situated within a larger structure. The second column gives the expected number of tests required to locate all positive samples in that group, using the optimal search algorithm. The third column gives the size of the two subgroups to be tested if the group tests positive. By following the division into successive subgroups, one can obtain the full binary structure that represents the optimal

TABLE 1: Successive divisions and expected number of tests

Size Expected Division	Size Expected Division	Size Expected Division
n **#tests**	***n*** **#tests**	***n*** **#tests**
3, 1.000699960, 1, 2	26, 1.017892624, 8, 18	49, 1.040072616, 15, 34
4, 1.001199900, 1, 3	27, 1.018792024, 8, 19	50, 1.041071609, 16, 34
5, 1.001699840, 2, 3	28, 1.019691384, 8, 20	51, 1.042070590, 17, 34
6, 1.002299730, 2, 4	29, 1.020590715, 8, 21	52, 1.043069521, 18, 34
7, 1.002899610, 2, 5	30, 1.021490155, 9, 21	53, 1.044068482, 19, 34
8, 1.003499490, 3, 5	31, 1.022389556, 10, 21	54, 1.045067394, 20, 34
9, 1.004199300, 3, 6	32, 1.023288926, 11, 21	**55, 1.046066265, 21,34**
10, 1.004899090, 3, 7	33, 1.024188296, 12, 21	56, 1.047164848, 21, 35
11, 1.005598870, 3, 8	**34, 1.025087627, 13, 21**	57, 1.048263370, 21, 36
12, 1.006298670, 4, 8	35, 1.026086758, 13, 22	58, 1.049361844, 21, 37
13, 1.006998450, 5, 8	36, 1.027085839, 13, 23	59, 1.050460296, 21, 38
14, 1.007798130, 5, 9	37, 1.028084881, 13, 24	60, 1.051558689, 21, 39
15, 1.008597780, 5, 10	38, 1.029083911, 13, 25	61, 1.052657102, 21, 40
16, 1.009397411, 5, 11	39, 1.030082893, 13, 26	62, 1.053755455, 21, 41
17, 1.010197051, 5, 12	40, 1.031081904, 13, 27	63, 1.054853758, 21, 42
18, 1.010996661, 5, 13	41, 1.032080865, 13, 28	64, 1.055952151, 21, 43
19, 1.011796321, 6, 13	42, 1.033079786, 13, 29	65, 1.057050485, 21, 44
20, 1.012595951, 7, 13	43, 1.034078807, 13, 30	66, 1.058148768, 21, 45
21, 1.013395561, 8, 13	44, 1.035077779, 13, 31	67, 1.059247031, 21, 46
22, 1.014295032, 8, 14	45, 1.036076710, 13, 32	68, 1.060345235, 21, 47
23, 1.015194462, 8, 15	46, 1.037075631, 13, 33	69, 1.061443638, 21, 48
24, 1.016093863, 8, 16	47, 1.038074503, 13, 34	70, 1.062541983, 21, 49
25, 1.016993263, 8, 17	48, 1.039073584, 14, 34	71, 1.063640278, 21, 50
89, 1.083409245, 34, 55	**144, 1.14925821, 55, 89**	**233, 1.26456020, 89, 144**
377, 1.46511596, 144, 233	**610, 1.81188758, 233, 377**	**987, 2.40799356, 377, 610**
1597, 3.426668, 610, 987	**2584, 5.1563753, 987, 1597**	**4181, 8.072368, 1597, 2584**
6765, 12.948090, 2584, 4181		

search. For example, if during the course of a procedure a group of size 233 occurs and tests positive, the subsequent subgroups should be of sizes 89 and 144; if the subgroup of 89 tests positive, further subgroups of 34 and 55 are tested and so on until subgroups testing negative are eliminated and only individual positive samples remain. When applied to a population of one million, the expected number of tests using this procedure is 1913.982. This compares favorably with the results of the above-mentioned student procedures that assume there are exactly 100 positive samples, rather than the probability $p = 0.0001$. (In an actual test situation, this would be adjusted slightly since 6765 does not divide exactly into one million. We will omit this refinement.)

The Fibonacci Conjecture

From the output of hundreds of computer runs with $0.9 < q < 1.0$ and n < 10,000, it appears that the optimal groupings are always the same, regardless of q. For instance, in an optimal structure, a test that includes 55 samples is always followed by tests that include 21 and 34 samples, and a test of 34 samples is followed by tests that include 13 and 21 samples. An inspection of Table 1 suggests a Fibonacci-based pattern.

Conjecture. If the optimal fixed structure calls for the inclusive test of a group of size n, then if n is the Fibonacci number F_k, the two subgroups tested within this group will be of sizes F_{k-1} and F_{k-2}. If n is not a Fibonacci number, then these two subgroups will be of sizes m and $(n-m)$, where at least one of these is a Fibonacci number, and there is exactly one Fibonacci number between m and $(n-m)$; there will be only one pair that satisfies the conditions.

Although the conjecture has been verified for 20 different values of q, it has not been proven.

We note that the running time to implement the Fibonacci conjecture is considerably shorter than the running time for the optimal search algorithm presented above.

Adjacent Remarks

Concerning the Fibonacci conjecture Neither a proof of the Fibonacci conjecture nor a counterexample has as yet emerged. There is a superficial resemblance to two results concerning binary searches: the "golden section search procedure" in nonlinear programming [**2**, pp. 179–183, **7**], and the study of Fibonacci trees in [**3**]. Yet they start from different assumptions and do not support any proof.

A broader definition of a valid search Definition 1 assumes a valid search procedure that is completely predetermined except for the omission of a test whose result is already known. This concise definition might be relaxed to accommodate the following common occurrence. Assume that we have reached a stage at which we are testing $\overline{ab}$, where a and b are substructures. If together they test positive and then a alone tests positive, we now know nothing about b. Indeed, we know nothing about any sample or group to the right of b so that, instead of proceeding with our specified program—which requires us to test b—it would be more economical to begin again with all samples about which nothing is known. That is, we would apply the algorithm of Theorem 4 to all samples to the right of a, creating a new structure, the number of whose tests we would add to those we already obtained. This would be done at every further occurrence of an $\overline{ab}$ situation. Under this procedure of constant restructuring, the expected number of tests for the above example would be reduced from 1913.982 to 1542.691. This result is comparable to those of other procedures applied to the more restrictive version of the problem that assumes exactly 100 positive samples [**5**].

It should be emphasized that in spite of such results, the algorithm of Theorem 4 has not been shown to be the optimal algorithm to apply in a constantly restructuring procedure.

An optimal procedure without the assumption of nesting While every study, including this one, has assumed that testing proceeds in a nested manner, this does not necessarily produce the optimal result. For a population as small as $n = 3$, if $q > \frac{1+\sqrt{33}}{8} \approx .843$, there is a nonnested procedure that gives a lower expected number of tests than the algorithm given above. However, the physical application of a nonnested algorithm might be impractical, as, for example, in searches within fixed circuitry.

Proofs

Several definitions of convenience will be made within the proofs, often using the sign "≡."

Proof of Theorem 2 By Theorem 1, the only test affected by the existence or nonexistence of the proposed test is the test immediately above it. Thus, we need consider only a structure such as $\overline{\overline{n_j}\dots\overline{\overline{n_k}\dots\overline{n_l}}\dots\overline{n_m}}$, where, in keeping with Theorem 1, we've noted the number of samples in groups j, k, l, and m. We assume that the groups are actually numbered successively so that $j < \cdots < k < \cdots < l < \cdots < m$, and a difference such as $m - j$ equals the exact number of groups between m and j plus one. The proposed, immediately disadvantageous test, includes groups k through l, and the test of all the groups from j through m is the next test outside it. There are two cases, given below.

Case 1. Group l is not group m; that is, there are groups between them; equivalently, $l - m > 0$. In this case, the test

$$\overline{\overline{n_j}\dots\overline{n_m}}\ \textit{without}\ \overline{\overline{n_k}\dots\overline{n_l}}\ \text{adds } 1 - q^{\sum_j^{m-1} n_i} - (m-j)q^{\sum_j^{m} n_i} \equiv E_2$$

where Σ indicates summation through the groups from left to right.

The test

$$\overline{\overline{n_j}\dots\overline{n_m}}\ \textit{with}\ \overline{\overline{n_k}\dots\overline{n_l}} \quad \text{adds} \quad 1 - q^{\sum_j^{m-1} n_i} - (m - j - (l-k))q^{\sum_j^{m} n_i} \equiv E_3,$$

$$\text{where } E_2 - E_3 = -(l-k)q^{\sum_j^{m} n_i} \equiv E_4 \quad \text{and} \quad E_4 < 0.$$

Since $\overline{\overline{n_k}\dots\overline{n_l}}$ was assumed to be disadvantageous, the total addition without $\overline{\overline{n_k}\dots\overline{n_l}}$ is better (lower) than with it.

Case 2. Group l is the same as group m; that is, $l - m = 0$. In this case, $\overline{\overline{n_j}\dots\overline{n_m}}$ *without* $\overline{\overline{n_k}\dots\overline{n_l}}$ is the same as in *Case 1*. $\overline{\overline{n_j}\dots\overline{n_m}}$ *with* $\overline{\overline{n_k}\dots\overline{n_l}}$ adds

$$1 - q^{\sum_j^{k-1} n_i} - (m - j - (m-k))q^{\sum_j^{m} n_i} = 1 - q^{\sum_j^{k-1} n_i} - (k-j)q^{\sum_j^{m} n_i} \equiv E_5 \text{ so that}$$

$$\begin{aligned} E_2 - E_5 &= q^{\sum_j^{k-1} n_i} + (k-j)q^{\sum_j^{m} n_i} - q^{\sum_j^{m-1} n_i} - (m-j)q^{\sum_j^{m} n_i} \\ &= q^{\sum_j^{k-1} n_i}\left[1 - q^{\sum_k^{m-1} n_i} - (m-k)q^{\sum_k^{m} n_i}\right] \equiv E_6, \end{aligned}$$

which, since $l = m$, is exactly a fraction (< 1) of the amount added by the disadvantageous test $\overline{\overline{n_k}\dots\overline{n_l}}$. But since the entire structure *with* $\overline{\overline{n_k}\dots\overline{n_l}}$ must also add on the positive value of $\overline{\overline{n_k}\dots\overline{n_l}}$, this is more than the fractional advantage it has over the structure *without* $\overline{\overline{n_k}\dots\overline{n_l}}$. ■

Two lemmas for the proof of Theorem 3

Lemma 1. *Given the structure* $\overline{\overline{n_1}\,\overline{n_2}\,\overline{n_3}\dots\overline{n_{k-1}}\,\overline{n_k}}$, *where* $n_k \geqslant n_1 \geqslant n_2 \geqslant n_3 \geqslant \cdots \geq n_{k-1}$, *and where the inclusive test is advantageous, then at least one of the following is better than the given structure:*

$$A = \overline{\overline{n_1}\,\overline{n_2}\,\overline{n_3}\dots\overline{\overline{\overline{n_{k-1}}\,\overline{n_k}}}} \ \textit{or}\ B = \overline{\overline{n_1}\,\overline{\overline{\overline{n_2}\,\overline{n_3}\dots\overline{n_{k-1}}\,\overline{n_k}}}}.$$

Proof. Above the k groups in the given structure, the addition of the overall (advantageous) test adds the negative value

$$1 - q^{n - n_k} - (k-1)q^n \equiv E_7.$$

For structure B, the addition of the inclusive test adds

$$[1 - q^{n-n_1-n_k} - (k-2)q^{n-n_1}] + [1 - q^{n_1} - q^n] \equiv E_8.$$

For structure A, the addition of the inclusive test adds

$$[1 - q^{n_{k-1}} - q^{n_{k-1}+n_k}] + [1 - q^{n-n_{k-1}-n_k} - (k-2)q^n] \equiv E_9.$$

Thus, we have to prove that either $E_8 - E_7 < 0$ or $E_9 - E_7 < 0$.

$$\begin{aligned}
E_8 - E_7 &= 1 - q^{n-n_1-n_k} - (k-2)q^{n-n_1} + 1 - q^{n_1} - q^n - 1 + q^{n-n_k} + (k-1)q^n \\
&= 1 - q^{n-n_1-n_k} - (k-2)q^{n-n_1} - q^{n_1} + q^{n-n_k} + (k-2)q^n \\
&= (1 - q^{n_1})(1 - q^{n-n_k-n_1} - (k-2)q^{n-n_1}) \equiv E_{10},
\end{aligned}$$

which is $(1 - q^{n_1})$ times the amount added by $\overline{\overline{n_2}\ldots\overline{n_k}}$.

Define $(1 - q^{n-n_k-n_1} - (k-2)q^{n-n_1})$, the amount added by $\overline{\overline{n_2}\ldots\overline{n_k}}$, as E_{11}. If E_{11} is negative, then E_8 is better than E_7, and the proof is finished. Assume, therefore, that E_{11} is positive, and consider $E_9 - E_7$ under this assumption.

$$\begin{aligned}
E_9 - E_7 &= 1 - q^{n_{k-1}} - q^{n_{k-1}+n_k} + 1 - q^{n-n_{k-1}-n_k} - (k-2)q^n - 1 + q^{n-n_k} + (k-1)q^n \\
&= 1 - q^{n_{k-1}} - q^{n_{k-1}+n_k} - q^{n-n_{k-1}-n_k} + q^{n-n_k} + q^n \\
&= 1 - q^{n_{k-1}} - q^{n_{k-1}+n_k} - q^{n-n_{k-1}-n_k}(1 - q^{n_{k-1}} - q^{n_{k-1}+n_k}) \\
&= (1 - q^{n-n_{k-1}-n_k})(1 - q^{n_{k-1}} - q^{n_{k-1}+n_k}) \equiv E_{12},
\end{aligned}$$

which is $(1 - q^{n-n_{k-1}-n_k})$ times the amount added by $\overline{\overline{n_{k-1}}\,\overline{n_k}}$.

Let

$$(1 - q^{n_{k-1}} - q^{n_{k-1}+n_k}) \equiv E_{13}.$$

If E_{13} is negative, then E_9 is better than E_7. Assume, therefore, that E_{13} is positive.

We will now show that the assumption that both E_{11} and E_{13} are positive leads to a contradiction, which will prove Lemma 1. Since E_7 has been assumed negative, to show that E_{11} is in fact negative, contrary to our assumption, we need only show that $E_7 - E_{11}$ is positive:

$$\begin{aligned}
E_7 - E_{11} &= 1 - q^{n-n_k} - (k-1)q^n - 1 + q^{n-n_k-n_1} + (k-2)q^{n-n_1} \\
&= q^{n-n_k-n_1} + (k-2)q^{n-n_1} - q^{n-n_k} - (k-1)q^n \\
&= q^{n-n_k-n_1}[1 + (k-2)q^{n_k} - q^{n_1} - (k-1)q^{n_1+n_k}] \\
&= q^{n-n_k-n_1}[1 - q^{n_1} - q^{n_k+n_1}] + q^{n-n_k-n_1}[(k-2)q^{n_k} - (k-2)q^{n_1+n_k}] \equiv E_{14}.
\end{aligned}$$

Now, we have assumed that $n_1 \geqslant n_{k-1}$, and therefore, the first quantity in brackets in the definition of E_{14} is at least as great as E_{13}. Since E_{13} was assumed positive and since the second bracketed quantity in the definition of E_{14} is positive, E_{14} must also be positive. The contradiction proves Lemma 1. ■

Lemma 2. *In Lemma 1, the structure* $A = \overline{\overline{n_1}\,\overline{n_2}\,\overline{n_3}\ldots\overline{\overline{n_{k-1}}\,\overline{n_k}}}$ *is always advantageous over the original structure* $\overline{\overline{n_1}\,\overline{n_2}\,\overline{n_3}\ldots\overline{n_{k-1}}\,\overline{n_k}}$.

Proof. Since Lemma 1 proved that at least one of A or B is always advantageous over the original $\overline{\overline{n_1}\,\overline{n_2}\,\overline{n_3}\ldots\overline{n_{k-1}}\,\overline{n_k}}$, we need only show that B advantageous implies A advantageous. This is true if E_{11} is negative implies that E_{12} is negative, too. Assuming that E_{13} is positive and remembering that $n_k \geqslant n_1 \geqslant \cdots \geqslant n_{k-1}$, we now compute the value of $\overline{n_1\,n_2 \ldots \overline{\overline{n_{j-1} \ldots n_k}}}$ *minus* the value of $\overline{n_1\,n_2 \ldots \overline{\overline{n_j \ldots n_k}}}$ This difference equals

$$\left[1 - q^{\sum_{j-1}^{k-1} n_i} - (k-j+1)q^{\sum_{j-1}^{k} n_i}\right] - \left[1 - q^{\sum_{j}^{k-1} n_i} - (k-j)q^{\sum_{j}^{k} n_i}\right]$$

$$= -q^{\sum_{j-1}^{k-1} n_i} - (k-j+1)q^{\sum_{j-1}^{k} n_i} + q^{\sum_{j}^{k-1} n_i} + (k-j)q^{\sum_{j}^{k} n_i}$$

$$= q^{\sum_{j}^{k-1} n_i}\left[-q^{n_{j-1}} - (k-j+1)q^{n_k+n_{j-1}} + 1 + (k-j)q^{n_k}\right]$$

$$= q^{\sum_{j}^{k-1} n_i}\left[1 - q^{n_{j-1}} - (k-j+1)q^{n_k+n_{j-1}} + (k-j)q^{n_k}\right] \equiv E_{15}.$$

The factor in brackets in E_{15} is $\geq [1 - q^{n_{j-1}} - (k-j+1)q^{n_k+n_{j-1}} + (k-j)q^{n_k+n_{j-1}}]$, which simplifies to $[1 - q^{n_{j-1}} - q^{n_k+n_{j-1}}]$, which itself is $\geq [1 - q^{n_{k-1}} - q^{n_k+n_{k-1}}]$ since $n_{j-1} \geqslant n_{k-1}$. This last term, $[1 - q^{n_{k-1}} - q^{n_k+n_{k-1}}]$, is exactly E_{13}, however, and therefore, we know that as the test is extended one more group to the left it grows more positive. Since E_{11} is eventually reached by such shifts to the left, it will be positive. Therefore, a negative E_{11} implies a negative E_{12}, and Lemma 2 is proved. ■

Rewriting and proof of Theorem 3 Three definitions are needed.

Definition 8. If a test has $n-1$ tests that include it, it is said to be on level n.

Definition 9. If a test includes s groups immediately beneath it, it is said to have an excess of $s-2$.

Definition 10. A test is said to have an excess if its excess is greater than or equal to 1.

We now rewrite Theorem 3 in a form that we can prove more conveniently.

Theorem 5. (rewritten) *If a structure contains any test with an excess, there is another structure on the same samples, none of whose tests has an excess, and that is at least as good as the given structure.*

Proof. The following is a method for changing the original structure into one without excess. At any stage, take a test A with excess, on the lowest possible level. Arrange the groups so that $n_k \geqslant n_1 \geqslant n_2 \geqslant n_3 \geqslant \cdots \geq n_{k-1}$. Since the value of A is $1 - q^{n-n_k} - (k-1)q^n$, by placing the largest group (maximum n_i) to the right, we can only be working to our advantage. Put a test over $\overline{n_{k-1}}\,\overline{n_k}$ so that $\overline{\overline{n_{k-1}\,n_k}}$ is now part of the structure. By Lemma 2, this is advantageous. If A is still negative, continue. The excess within A will either be eaten away in this manner, or, at some point, A will turn positive. If this happens, by Theorem 2, we can erase the test A without disadvantage, and the remaining excess is lifted to a higher level. In fact, if the next test above A is made positive by the removal of A, by Theorem 2, we can drop that test also, and the excess will go to an even higher level. In this way, all the excesses are eliminated. ■

Proof of Theorem 4, the optimal algorithm. We see immediately why we need only go up to $\lceil n/2 \rceil$, for the possibilities in A are independent of the order of n' and $n - n'$, and those in B are more advantageous if the smaller group is placed to the left. That the procedure is optimal we see as follows: Consider $O_q(n)$, the optimal structure on n samples. If it has no inclusive test, then it is the sum of two smaller structures (one or both of which might again have no inclusive test.) The minimum of the possibilities of type A will clearly give us this. If $O_q(n)$ has an inclusive test, by Theorem 3, the test cannot include more than two substructures within it so that all permissible cases are included in B. That we always want to use $O_q(n')$ and $O_q(n-n')$ can be seen from the fact that, since the test over all n samples is independent of what exists within the next layer inside, we would certainly never want to substitute for a fully tested $O_q(n')$ or $O_q(n-n')$ another fully tested structure on n′ samples or $(n-n')$ samples. Furthermore, by Theorem 2, if $O_q(n')$ or $O_q(n-n')$ were not fully tested, we would never want to add a disadvantageous test for the purpose of gaining an advantage with a test of all n samples. ■

REFERENCES

1. R. Dorfman, The detection of defective members of large populations, *Ann. of Math. Stat.* **14** (1943), 436–440.
2. D-Z. Du, F. Hwang, *Combinatorial Group Testing and Its Applications*. World Scientific, Singapore, 1993.
3. Y. Horibe, An entropy view of Fibonacci trees, *Fibonacci Quart.* **20** (1982) 168–178.
4. F. K. Hwang, Y. C. Yao, Cutoff point and monotonicity properties for multinomial group testing, *SIAM J. Discrete Math.* **2** no. 4 (1989) 500–507.
5. W. Jarnicki, M. Schweizer, S. Wagon, Follow-up on disease detection, *College Math. J.* **38** no. 2 (2007) 134.
6. E. Knill, W. Bruno, D. Torney, Non-adaptive group testing in the presence of errors, *Discrete Appl. Math.* **88** (1998) 261–290.
7. D. G. Luenberger, *Linear and Nonlinear Programming*. Addison Wesley, Reading, 1984. 197–200.
8. M. Sobel, P.A. Groll, Group testing to eliminate efficiently all defectives in a binomial sample, *Bell System Tech. J.* **28** (1959) 1179–1252.
9. D. J. Teague, Group testing: Four student solutions to a classic optimization problem, *College Math. J.* **37** no. 4 (2006) 261–268.
10. P. Ungar, The cutoff point for group testing, *Comm. Pure. Appl. Math.* **13** (1960) 49–54.
11. S. Zimmerman, A testing algorithm, IBM Research, Yorktown Heights, NY, RC 2838 (April 1, 1970).

Summary. The use of group testing to locate all instances of disease in a large population of blood samples was first considered more than 70 years ago. Since then, several procedures have been used to lower the expected number of tests required. The algorithm presented here, in contrast to previous ones, takes a constructive rather than a top-down approach. As far as could be verified, it offers the first proven solution to the problem of finding a predetermined procedure that guarantees the minimum expected number of tests. Computer results strongly suggest that the algorithm has a Fibonacci-based pattern.

SETH ZIMMERMAN (MR Author ID: 571321) was a professor of mathematics in the Bay Area and a humanities lecturer in the U.K. His areas of research are probability, cosmology, number theory, entropy, and combinatorial geometry. A parallel literary career includes poetry, fiction, and translations of Dante and Mandelstam.

Real Mathematics Fact

Snapple includes “Real Facts” on the underside of the caps to its beverages. A number of them have been about mathematics, including the following.

Snapple’s “Real Fact” #1326 states that, “The horizontal line between two numbers in a fraction is called a vinculum.” Others may know the line that separates the numerator from the denominator as the “fraction bar.”

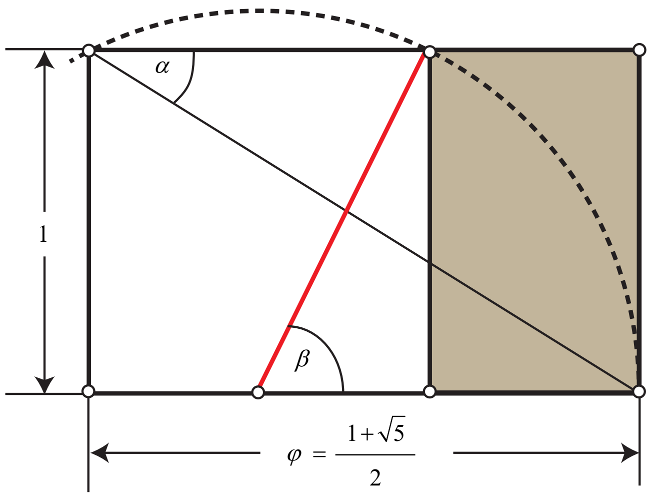
—Submitted by Nick Roopas,
Ann Arbor, MI

Proof Without Words: Arctangent of Two and the Golden Ratio

ÁNGEL PLAZA
Universidad de Las Palmas de Gran Canaria
Las Palmas, Canaria, Spain
angel.plaza@ulpgc.es

Theorem. Let φ be the golden ratio, $\varphi = \dfrac{1+\sqrt{5}}{2}$. Then $\arctan\left(\dfrac{1}{\varphi}\right) = \dfrac{\arctan 2}{2}$.

Proof.



$$\alpha = \arctan\left(\frac{1}{\varphi}\right), \quad \beta = 2\alpha = \arctan 2. \qquad \blacksquare$$

Summary. It is proved without words that the golden ratio, φ, and the arctangent of 2 are related.

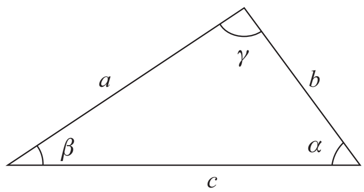
ÁNGEL PLAZA (MR Author ID: 350023) received his masters degree from Universidad Complutense de Madrid in 1984 and his Ph.D. from Universidad de Las Palmas de Gran Canaria in 1993, where he is a full professor in applied mathematics. He is interested in mesh generation and refinement, combinatorics and visualization support in teaching and learning mathematics.

Math. Mag. **90** (2017) 179. doi:10.4169/math.mag.90.3.179.
MSC: Primary 51M04

A Proof of the Law of Sines Using the Law of Cosines

PATRIK NYSTEDT
University West
Trollhättan, Sweden
patrik.nystedt@hv.se

Two of the most well known relations between the sides and the angles in a triangle are, respectively, the law of sines (LS)



$$\frac{\sin\alpha}{a} = \frac{\sin\beta}{b} = \frac{\sin\gamma}{c} \tag{1}$$

and the law of cosines (LC)

$$b^2 + c^2 - 2bc\cos\alpha = a^2, \tag{2}$$

$$a^2 + c^2 - 2ac\cos\beta = b^2, \tag{3}$$

$$a^2 + b^2 - 2ab\cos\gamma = c^2. \tag{4}$$

Traditionally, (LS) and (LC) are proved separately (see, e.g., [**1**]). What seems to be less known is that it is possible to deduce one of the laws from the other. Indeed, [**3**] demonstrates that it is possible to prove (LC) from (LS). Conversely, Kirschen and Serulneck [**2**] show that (LC) implies (LS). The aim of this note is to give a proof of the latter statement, which is similar, but different and shorter, than the proof used in [**2**]. Namely, suppose that (2), (3), and (4) hold. Since $\sin\alpha$, $\sin\beta$, and $\sin\gamma$ are positive, it is enough to show the squared version of (1). Using the Pythagorean trigonometric identity $\sin^2\alpha + \cos^2\alpha = 1$, equation (2) and the difference of squares factorization technique, we get that

$$\frac{\sin^2\alpha}{a^2} = \frac{1-\cos^2\alpha}{a^2} = \frac{4b^2c^2 - 4b^2c^2\cos^2\alpha}{4a^2b^2c^2} = \frac{4b^2c^2 - (b^2+c^2-a^2)^2}{4a^2b^2c^2}$$

$$= \frac{(a-b+c)(a+b-c)(b+c-a)(b+c+a)}{4a^2b^2c^2}. \tag{5}$$

The rational expression in (5) is symmetric in the variables a, b, and c. In particular, it is invariant under the cyclic permutation $p : (a,b,c) \mapsto (c,a,b)$. Thus, we can deduce that both $(\sin^2\beta)/b^2$ and $(\sin^2\gamma)/c^2$ equal the rational expression in (5), by applying p twice.

Math. Mag. **90** (2017) 180–181. doi:10.4169/math.mag.90.3.180. © Mathematical Association of America
MSC: Primary 51M04, Secondary 97G60

Acknowledgment I wish to thank the anonymous referees for their careful reading of my manuscript and their many comments and suggestions.

REFERENCES

1. I. M. Gelfand, M. Saul, *Trigonometry.* Springer, Birkhäuser, Boston, 2001.
2. L. S. Kirschen, J. A. Serulneck, The law of sines from the law of cosines, *The Mathematics Teacher*, NCTM **88** no. 1 (January 1995) 76.
3. Wikipedia contributors, Law of cosines, *Wikipedia, The Free Encyclopedia*, `http://en.wikipedia.org/wiki/Law_of_cosines`.

Summary. We give a proof of the law of sines using the law of cosines.

PATRIK NYSTEDT (MR author ID: 641289) holds a permanent appointment as professor of mathematics in the department of engineering science at University West in Trollhättan, Sweden. Dr. Nystedt's research interests include category theory and noncommutative algebra. He received his B.S. and Ph.D. degrees in pure mathematics from Gothenburg University, Sweden.

On the February and April Pinemi puzzles

We thank eagle-eyed reader Rob Pratt for his correspondence regarding the February and April Pinemi puzzles. Rob noticed that there were two solutions to the February puzzle as the || and ||| in positions (8, 6), (8, 7), (9, 6), and (9, 7) can be reversed.

He also recognized that there were two mistakes in the April puzzle. The 4 in position (1, 9) should be a 3. And the 6 in position (9, 4) should be a 7. The submitted puzzle for April was correct, as the two errors were introduced in the production process. The correct values appear in the solution. We apologize to the solvers of this puzzle and to the puzzle creator, Lai Van Duc Thinh.

SOLUTION TO PINEMI PUZZLE

<table>
<tr><td>|</td><td>5</td><td>|</td><td>5</td><td>|</td><td>||</td><td>7</td><td>||</td><td>|</td><td>||</td></tr>
<tr><td>8</td><td>|||</td><td>9</td><td>|||</td><td>9</td><td>8</td><td>|||</td><td>9</td><td>|</td><td>5</td></tr>
<tr><td>|||</td><td>|</td><td>10</td><td>|</td><td>||</td><td>7</td><td>6</td><td>||</td><td>8</td><td>|</td></tr>
<tr><td>||</td><td>10</td><td>||</td><td>9</td><td>|</td><td>|</td><td>7</td><td>8</td><td>|</td><td>|||</td></tr>
<tr><td>3</td><td>|</td><td>|</td><td>|</td><td>|</td><td>|</td><td>|||</td><td>8</td><td>||</td><td>8</td></tr>
<tr><td>6</td><td>9</td><td>||</td><td>||</td><td>10</td><td>||</td><td>10</td><td>||</td><td>7</td><td>||</td></tr>
<tr><td>|||</td><td>||</td><td>10</td><td>11</td><td>|||</td><td>9</td><td>|</td><td>|</td><td>8</td><td>5</td></tr>
<tr><td>6</td><td>10</td><td>||</td><td>||</td><td>10</td><td>||</td><td>|</td><td>8</td><td>||</td><td>|</td></tr>
<tr><td>|</td><td>8</td><td>||</td><td>9</td><td>|</td><td>||</td><td>10</td><td>|||</td><td>10</td><td>|</td></tr>
<tr><td>4</td><td>|||</td><td>6</td><td>|</td><td>|</td><td>6</td><td>||</td><td>8</td><td>|||</td><td>4</td></tr>
</table>

Extending Two Classic Proofs of the Pythagorean Theorem to the Law of Cosines

KEVIN K. FERLAND
Bloomsburg University
Bloomsburg, PA 17815
kferland@bloomu.edu

Even though it is well known that the Pythagorean theorem is a special case of the law of cosines, it is less obvious whether specific proofs of the Pythagorean theorem can be extended to prove the law of cosines. In this paper, we consider two geometric dissection proofs of the Pythagorean theorem and explain how to modify these original proofs for general triangles.

Throughout this paper, let T denote a triangle with side lengths a, b, and c, and let C be the measure of the angle opposite the side of length c. Recall that the law of cosines asserts that

$$a^2 + b^2 - 2ab\cos C = c^2. \tag{1}$$

Classic proofs of the Pythagorean theorem

The Pythagorean theorem deals with the case in which T is a right triangle, namely $C = 90^\circ$ and $\cos C = 0$ in Equation (1). Two well-known proofs, classified as algebraic proofs in [**11**], are accomplished using Figure 1, in which the pictured points O mark the centers of the displayed squares for future reference.

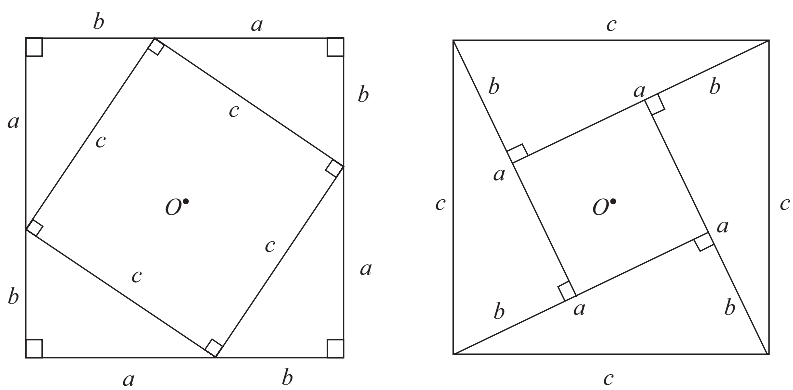


Figure 1 Area of big square = area of little square + area of four triangles.

In each picture in Figure 1, we see that the area of the outside square equals the area of inside square plus the areas of the four triangles.

In Proof 1, the equivalent areas on the left-hand side of Figure 1 yield the equations

$$(a+b)^2 = c^2 + 4\left(\frac{1}{2}ab\right), \tag{2}$$

$$a^2 + 2ab + b^2 = c^2 + 2ab.$$

Math. Mag. **90** (2017) 182–186. doi:10.4169/math.mag.90.3.182.
MSC: Primary 51M04

Canceling $2ab$ from both sides gives us the equation of the Pythagorean theorem,

$$a^2 + b^2 = c^2.$$

Similarly, Proof 2 is obtained by considering the equivalent areas shown on the right-hand side of Figure 1.

Working toward the law of cosines

In [**10**], several standard proofs of the law of cosines are presented, such as those using the distance formula or trigonometry. Trigonometric arguments are also given in [**7**] and [**4**], and other simple geometric proofs are presented in [**1**], [**3**], and [**6**]. Proofs of particular interest here that consider areas inside of wisely chosen n-gons are given in [**2**], [**5**], [**8**], and [**9**]. However, each of those only considers special cases in which either $360^\circ = (180^\circ - C)n$ or $360^\circ = Cn$, for some integer $n \geq 3$. The area-based arguments given there are generalized here and exploited to full effect to obtain Equation (1) for all C.

To allow for our generalization of Proof 1, we make a minor reinterpretation of this classic proof. We use instead Figure 2, which is seen as a quarter of the original picture, with additional lines joining the corners of the squares to the center O. There, the area of the dark triangle $\triangle KMO$ equals the area of the dotted triangle $\triangle LNO$ plus the area of triangle T. The resulting change is to simply replace Equation (2) with a quarter of itself,

$$\frac{1}{4}(a+b)^2 = \frac{1}{4}c^2 + \frac{1}{2}ab,$$

and otherwise the proof is unchanged. The key observation is that $\triangle KLO$ is congruent to $\triangle MNO$.

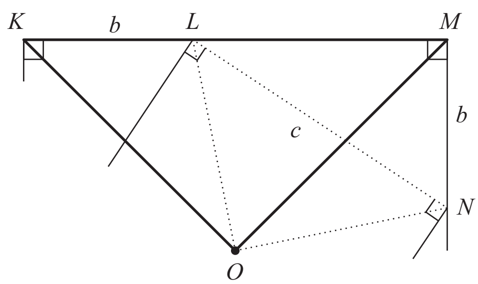


Figure 2 Area of dark $\triangle KMO$ = area of dotted $\triangle LNO$ + area of triangle T.

Extending the classic dissection proofs

To extend the classic proofs beyond right triangles, we will invoke some elementary results from trigonometry. First, observe that the area of our triangle T is $\frac{1}{2}ab\sin C$. This is easily seen by regarding a as the base length and observing that the height is then given by $b\sin(180^\circ - C) = b\sin C$. The real workhorse for our extended dissection proofs will be the following result on isosceles triangles.

Lemma. The area of an isosceles triangle with base length s and base angles of measure θ is $\frac{s^2}{4}\tan\theta$; see Figure 3.

Proof. Since $h = \frac{s}{2}\tan\theta$, the area is $\frac{1}{2}sh = \frac{s^2}{4}\tan\theta$. ■

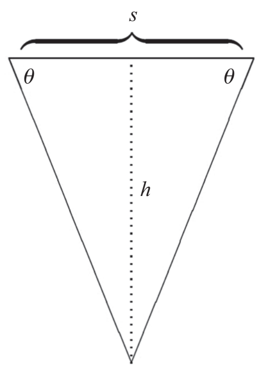


Figure 3 Isosceles triangle with base length s, base angles of measure θ, and height h.

In light of the expression in our lemma, we will also invoke the identity

$$\tan\theta = \frac{\sin 2\theta}{1+\cos 2\theta},$$

that we get from $(1+\cos 2\theta)\tan\theta = (2\cos^2\theta)\tan\theta = 2\cos\theta\sin\theta = \sin 2\theta$.

Our extension of Proof 1 is now based on Figure 4, in which $\triangle JKL$ and $\triangle LMN$, each congruent to T, have been placed so that K, L, and M are collinear.

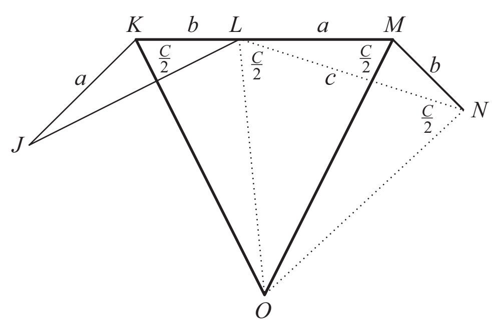


Figure 4 Area of dark $\triangle KMO$ = area of dotted $\triangle LNO$ + area of triangle T.

In Figure 4, point O is picked so that $\overrightarrow{KO}$ bisects $\angle JKL$ and $\overrightarrow{MO}$ bisects $\angle LMN$. Now, both $\triangle KMO$ and $\triangle LNO$ are isosceles triangles with base angles of measure $\frac{C}{2}$. Since $\triangle KLO$ is congruent to $\triangle MNO$, we see in Figure 4 that the area of dark triangle $\triangle KMO$ equals the area of dotted triangle $\triangle LNO$ plus the area of triangle T. Those equivalent areas and our lemma yield the equation

$$\frac{(a+b)^2}{4}\tan\frac{C}{2} = \frac{c^2}{4}\tan\frac{C}{2} + \frac{1}{2}ab\sin C,$$

which upon application of our tangent identity becomes

$$\frac{(a+b)^2}{4}\frac{\sin C}{1+\cos C}=\frac{c^2}{4}\frac{\sin C}{1+\cos C}+\frac{ab}{2}\sin C.$$

Multiplying this by $4(1+\cos C)/\sin C$ now yields

$$(a+b)^2=c^2+2ab(1+\cos C),$$

which further simplifies to give the law of cosines, as desired.

To extend Proof 2, we first review that proof here in greater generality. If $360^\circ = Cn$ for some $n \geq 3$, then we can appeal to Figure 5, which shows the case in which $n = 6$. There we assume that $a \geq b$ and note that the right-hand side of Figure 1 is the case in

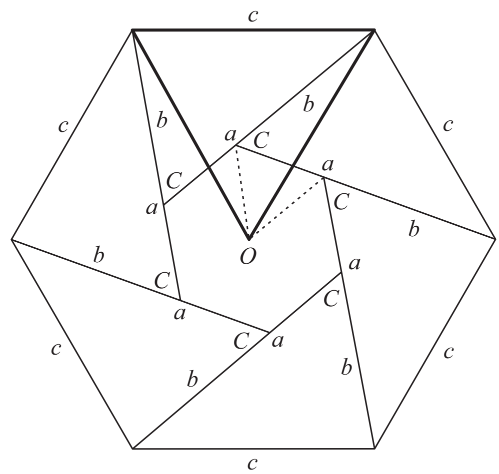


Figure 5 Area of big n-gon = area of little n-gon + area of n triangles.

which $n = 4$. In Figure 5, we see that the area of the outside n-gon equals the area of the inside n-gon plus the areas of n copies of triangle T. Those equivalent areas and our lemma yield the equations

$$n\frac{c^2}{4}\tan\left(\frac{180^\circ - C}{2}\right)=n\frac{(a-b)^2}{4}\tan\left(\frac{180^\circ - C}{2}\right)+n\frac{1}{2}ab\sin C,$$

$$\frac{c^2}{4}\tan\left(\frac{180^\circ - C}{2}\right)=\frac{(a-b)^2}{4}\tan\left(\frac{180^\circ - C}{2}\right)+\frac{1}{2}ab\sin C \qquad (3)$$

Multiplying Equation (3) by 4 and applying our tangent identity then gives

$$c^2\frac{\sin(180^\circ - C)}{1+\cos(180^\circ - C)}=(a-b)^2\frac{\sin(180^\circ - C)}{1+\cos(180^\circ - C)}+2ab\sin C.$$

Noting that $\sin(180^\circ - C) = \sin C$ and $\cos(180^\circ - C) = -\cos C$, further algebraic simplification now easily yields again the law of cosines.

Conclusions

In our extension of Proof 1, if $360^\circ = (180^\circ - C)n$ for some $n \geq 3$, then n of the dark triangles from Figure 4 form a regular n-gon, as do n of the dotted triangles, and our proof directly follows the type of argument reflected in Figure 1, in which $n = 4$. The cases in which $n = 3$ or 6 are presented in [**5**], and $n = 5$ is also addressed there but with a more complicated area consideration. In [**2**], the case in which $n = 3$ is handled, while the argument for all n is given in [**8**] and reiterated in [**9**]. Further, in [**8**], it is wondered if their argument can be generalized for angle measures that are rational multiples of 360°. Of course, here we have given an argument for all C.

In our extension of Proof 2, note that the cases in which $n = 6$ or 3 are considered in [**5**]. As in our extension of Proof 1, the restriction that $360^\circ = Cn$ can be lifted by focusing only on the dark triangle (and the dotted triangle) in Figure 5 and starting with Equation (3).

Acknowledgment The author gratefully acknowledges valuable improvements suggested by the reviewers and the editor.

REFERENCES

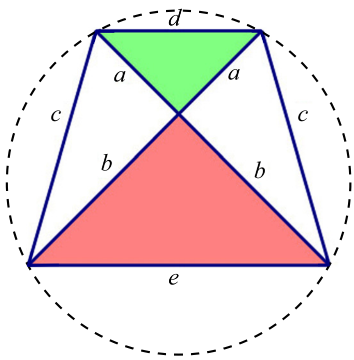
1. M. D. Edwards, A possibly new proof of the law of cosines, *Amer. Math. Monthly* **121** no. 2 (2014) 149.
2. J. Kovacevic, D. Veljan, A triangular analogue of the pythagorean theorem, *The Mathematical Gazette* **80** (1996) 550–554.
3. D. McConnell, *The Illustrated Law of Cosines*, 2010, http://www.cut-the-knot.org/pythagoras/DonMcConnell.shtml.
4. J. Molokach, Law of cosines—A proof without words, *Amer. Math. Monthly* **121** no. 8 (2014) 722.
5. B. Polster, M. Ross, One-glance(ish) proofs of Pythagoras' theorem for 60-degree and 120-degree triangles, *Math. Mag.* **89** no. 1 (2016) 47–54.
6. T. A. Sipka, Proof without words: Law of cosines for $\theta < \pi/2$, *Math. Mag.* **61** no. 4 (1988) 259.
7. S. L. Thompson, Classroom notes: Note on the law of cosines, *Amer. Math. Monthly* **58** no. 10 (1951) 698–699.
8. D. Veljan, An analogue of the Pythagorean theorem with regular n-gons instead of squares, *Elem. Math.* **51** no. 4 (1996) 156–160.
9. ———, The 2500-year-old Pythagorean theorem, *Math. Mag.* **73** no. 4 (2000) 259–272.
10. Wikipedia, *Law of cosines*, 2016, https://en.wikipedia.org/wiki/Law_of_cosines.
11. ———, *Pythagorean theorem*, 2016, https://en.wikipedia.org/wiki/Pythagorean_theorem.

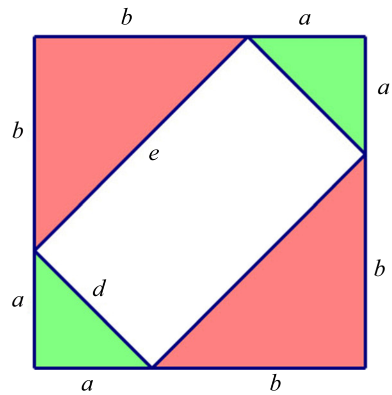
Summary. Two well-known proofs of the Pythagorean theorem are generalized to prove the law of cosines in a geometrically elegant way by computing areas.

KEVIN K. FERLAND (MR Author ID: 343376) received his Ph.D. in mathematics from Syracuse University in 1999 and is currently a professor of mathematics at Bloomsburg University. He is the author of the textbook *Discrete Mathematics: An Introduction to Proofs and Combinatorics* and has published several papers on the toughness of graphs. Kevin also enjoys both pursuing the combinatorics of and doing crossword puzzles.

Proof Without Words : Pythagorean Theorem via Ptolemy's Theorem

NAM GU HEO
Korea National University of Education
Cheongju, Chungcheongbukdo, Republic of Korea
mimirul@nate.com





$$(a+b)^2 = c^2 + de \qquad (a+b)^2 = a^2 + b^2 + de$$

$$\Rightarrow \; c^2 = a^2 + b^2$$

Ptolemy's Theorem For a convex quadrilaterial *ABCD* inscribed in a circle, the sum of the products of the two pairs of opposite sides equals the product of its two diagonals.

Summary. We provide a visual proof of Pythagorean theorem. The main idea of the proof is to compute $(a+b)^2$ in two different ways: one with aid of Ptolemy's theorem and the other one by dissecting a square.

NAM GU HEO (MR Author ID: 1106949) is a math teacher at Songchon high school and a part-time lecturer at Korea National University of Education.

Math. Mag. **90** (2017) 187. doi:10.4169/math.mag.90.3.187. © Mathematical Association of America
MSC: Primary 51M04

Serving Strategy in Tennis: Accuracy versus Power

YIGAL GERCHAK
Tel-Aviv University
Tel-Aviv 69978 ISRAEL
yigal@post.tau.ac.il

D. MARC KILGOUR
Wilfrid Laurier University
Waterloo, ON N2L 3C5 CANADA
mkilgour@wlu.ca

Tennis is unusual, perhaps unique, among sports in that the initiator of a sequence of play is allowed one free second chance should the first attempt fail. Competition for each point begins with a serve, and the server may serve again—without penalty—if the first serve faults, i.e., does not land in play. (In-bounds serves that touch the net, called "lets," are not considered to be faults and are always re-served.) We consider the strategic implications of this "free second chance," asking what sort of serve takes maximum advantage of this opportunity, whether and how the second serve should differ from the first and how the answers to these questions depend on the functions or parameters that describe the play.

The more powerful a serve, the more difficult it is to return, but the greater the probability that it faults, i.e., does not land in play. Thus, there is a tradeoff: A powerful, or "fast," serve is more likely to fault but more likely to win the point if it does not fault. Empirically, the opportunity to take a second serve affects strategy in tournament tennis, in that most players try a powerful but risky first serve, and follow with a softer but more accurate second serve if the first faults. The prevalence of these choices can be inferred from the statistics that, in men's professional tennis, about 60% of first serves and about 95% of second serves land in play. Of the first serves that do not fault, about 75% win the point; of the second serves, much closer to 50%. (The first three percentages are slightly lower for women's professional tennis, but the proportions are about the same.)

Observers have often asked whether a player's second serve should be more powerful than is typical, perhaps even as powerful as the first serve. At present, no professional tennis player regularly hits "two first serves," although Pete Sampras and a few others have done so occasionally, perhaps in part to surprise the opponent. More generally, what are the best strategies for these serves, and how different should they be?

For example, a recent New York Times article [**3**] makes the case for a second serve that is similar to the first, attributing current practice to a psychological phenomenon that amounts to a preference for losing late rather than early. In other words, the argument is that tennis players act as if they prefer losing in a rally to double-faulting a serve. The alternative viewpoint is that "safer" second serves are a form of insurance against the failure of high-speed first serves. These claims stimulated an interesting debate in the economists' blog Cheap Talk [**1, 2**], the latter contribution referring back to a model in [**5**]. The relatively few operations research analyses of strategy in tennis, including [**4, 8, 9**], do not settle this question. But the latter comes close, classifying serves as fast or slow and asking whether the optimal speed category of a single serve should depend on the current score of the match.

Math. Mag. **90** (2017) 188–196. doi:10.4169/math.mag.90.3.188. © Mathematical Association of America
MSC: Primary 90C99, Secondary 91B06; 60A99

Our models originate with the stylized model of David Gale [**6**]—see also Gilman [**7**]. The server selects a serving strategy on a continuum of levels of power or its surrogate, risk of fault. If the first serve faults, the strategy for the second serve—selected at the same time as the first-serve strategy—is implemented. For each serve, the conditional probability of winning the point (immediately or eventually) given that the serve does not fault is assumed known and is increasing in the probability of fault. Thus, if you run a greater risk of faulting, you have a higher probability of winning immediately, but only if you don't fault. This is the tradeoff of power and accuracy.

In a two-stage model, of course, one can solve for the optimal second-serve strategy and then use it to find the best first serve. Gale's elegant geometric argument shows that the second serve should be weaker than the first. Our analytic approach not only confirms this conclusion but also permits parametric representations and numerical measures of the relative advantage. We provide examples with explicit solutions and then extend the model so that it includes the rally that follows the successful return of a serve. Thus, we draw conclusions about how a player's optimal serving strategy reflects his or her ability in a rally.

Basic model

The server's decision problem is to choose strategies for two serves, where the second serving strategy is to be implemented only if the first serve faults. The choice of serve strategy is a tradeoff, usually described as "power versus accuracy." We model this tradeoff by thinking of the server's strategy as the choice of the probability, say x, that the serve will not fault. Functionally related to the nonfault probability is the conditional probability, $f(x)$, that the server will win the point given that the serve does not fault. (The server may win because (1) the receiver is unable to return the serve, or (2) the receiver makes the return but the server wins the subsequent rally. The probabilities of these two events are combined in $f(x)$; later, we present a model that separates them.)

We assume that $f(x)$, the *conditional win probability*, is a differentiable, strictly decreasing function satisfying $0 \leq f(x) \leq 1$. Thus, a powerful serve corresponds to a low value of x (likely to fault) and a high value of $f(x)$ (the serve is likely to be unreturnable, provided it does not fault). Similarly, a softer serve is unlikely to fault (x is high) but is more likely to be returned if it does not fault (and therefore the server is less likely to win the point, so $f(x)$ is lower).

Suppose that the first serve has faulted, and consider the second serve. If the server's second-serve strategy is denoted p, the server wins the point with probability

$$W(p) = pf(p).$$

Observe that $W(p)$ is concave if and only if $pf''(p) < -2f'(p)$ for all $p \in (0, 1]$); in particular, if $f(p)$ is concave, then so is $W(p)$. Henceforth, we assume that $W(p)$ is concave, and therefore that $W(p)$ has a unique maximum in [0, 1], at say $p = p^*$. Differentiating $W(p)$ and equating to zero produces

$$f(p^*) + p^* f'(p^*) = 0. \tag{1}$$

There are two cases. By the intermediate value theorem, if $f(1) < -f'(1)$, then there is a unique value of $p \in [0, 1)$ that solves (1); this value must be $p = p^*$. Otherwise, $f(1) \geq -f'(1)$, and (1) has no solution satisfying $p < 1$ so that $W(p)$ is maximized by $p = p^* = 1$. Note that, in either case, $p^* > 0$, $W(p^*) = p^* f(p^*) > 0$, and $p^* = 1$ if and only if $f(1) \geq -f'(1)$.

Now, we return to the first serve and assume that, if it faults, an optimal second serve characterized by $p = p^*$ will follow. If nonfault probability of the first serve is q, then the server wins the point with probability

$$V(q) = qf(q) + (1-q)W(p^*) = qf(q) + (1-q)p^*f(p^*).$$

Since W is concave, V is also. Therefore, $V(q)$ has a unique maximum on [0, 1], say at $q = q^*$. Differentiating $V(q)$ and equating to zero produces

$$f(q) + qf'(q) = p^*f(p^*). \tag{2}$$

If (2) has a solution in [0, 1), it must be unique and is therefore $q = q^* < 1$. If (2) has no solution in [0, 1), then $V(q)$ is a maximum at $q^* = 1$. In either case, the optimal first-serve strategy satisfies $q^* > 0$ and $q^* = 1$ if and only if $f(1) \geq -f'(1) + p^*f(p^*)$.

Gale's theorem shows, by a geometric argument, that an optimal first serve will be faster (riskier) than the second. We provide a new proof.

Proposition. *(Gale, [**6**]) $q^* \leq p^*$, with equality if and only if $q^* = 1$.*

Proof. Because $p^*f(p^*) > 0$, it follows that $f(1) \geq -f'(1) + p^*f(p^*)$ implies $f(1) \geq -f'(1)$. Therefore, $q^* = 1$ implies $p^* = 1$. If $-f'(1) + p^*f(p^*) > f(1) \geq -f'(1)$, then $q^* < 1$ and $p^* = 1$. Finally, if $f(1) < -f'(1)$, then $q^* \in (0, 1)$ is the solution of (2) and $p^* \in (0, 1)$ is the solution of (1). Both equations have the same left side, $f(x) + xf'(x)$, which is a strictly decreasing function. Since the right side of (2) is positive and the right side of (1) is zero, we must have $q^* < p^*$. ■

Thus, the first serve is always at least as risky as the second and strictly riskier unless both the first serve and the second serve are chosen with the absolute minimum of risk. This is general support for the softer second-serve strategy that is close to universal in tournament play.

But by how much should the optimal second serve differ from the optimal first serve? In other words, how great is the gap between q^* and p^*? To answer this question, we need a more specific model of the conditional win probability, $f(\cdot)$. We proceed first to address this gap in the context of two plausible forms for the conditional win probability $f(\cdot)$. Then we will build a richer model of the probability of winning the point by explicitly modeling the events after the serve—the return and the rally.

Example. Power Conditional Win Probability

Perhaps the simplest model for the conditional win probability is $f(x) = u - vx^w$, provided the values of the parameters u, v, and w satisfy $w > 0$ and $0 < v \leq u \leq 1$, which guarantee that $f(x)$ is strictly decreasing and that $0 \leq f(x) \leq 1$. We also impose the additional assumption that $u < v(1 + w)$, which ensures that p^*, as determined by (1), satisfies $0 < p^* < 1$. Because

$$W''(x) = xf''(x) + 2f'(x) = -wv(1+w)x^{w-1} < 0,$$

$W(\cdot)$ is concave. Moreover, $f(1) < -f'(1)$. The direct solution of (1) and (2) is now

$$p^* = \left[\frac{u}{v(1+w)}\right]^{\frac{1}{w}},$$

$$q^* = \left[\frac{u}{v(1+w)}\left(1 - \left(\frac{w}{1+w}\right)\left[\frac{u}{v(1+w)}\right]^{\frac{1}{w}}\right)\right]^{\frac{1}{w}}.$$

TABLE 1: Power conditional win probability: optimal strategies for various values of w at $u = v = 1$

w	p^*	q^*	$\frac{p^*-q^*}{p^*}$
$\frac{1}{2}$	0.4444	0.3225	0.2743
1	0.5	0.375	0.25
2	0.5774	0.4528	0.2157

How different are p^* and q^*? A convenient unitless measure is the proportionate increase in fault probability of the first serve relative to the second. In this example, it equals

$$\frac{p^* - q^*}{p^*} = 1 - \left(1 - \left(\frac{w}{1+w}\right)\left[\frac{u}{v(1+w)}\right]^{\frac{1}{w}}\right)^{\frac{1}{w}}.$$

It is easy to see that, if $w = 1$, this expression equals $\frac{u}{4v}$; for example, if $u = v = 1$, the probability of a fault on the first serve is optimally 25% greater than on the second. When $w = \frac{1}{2}$, the increase in fault probability becomes $\frac{8u^2}{27v^2}\left[1 - \frac{2u^2}{27v^2}\right]$; when $u = v = 1$, this gap in fault probabilities equals $\frac{200}{729} \approx 27.4\%$.

In general, it can be shown that, if $w > 1$, the value we obtained for $w = 1$ is an upper bound for the gap in fault probabilities. Table 1 confirms these values for $w = \frac{1}{2}$ and $w = 2$ in the case $u = v = 1$.

Extended model

We now develop a more nuanced model of how a serve can result in the server winning a point. If the serve does not fault, the server wins the point if either the receiver (opponent) is unable to return the serve, or the server wins the rally that follows the return. The probabilities associated with the latter two possibilities may have different effects on the optimal serving strategies, p^* and q^*.

As previously, we model the serving strategy by the probability, q, that the serve does not fault. But now we propose a model of a rally (see below) and a function that measures the probability that the serve leads to a rally because the receiver returns the serve. We assume a probability, $g(q)$, that the receiver cannot return the serve, given that it does not fault. We assume that $g(q)$, the *no-return probability* is a continuous, decreasing function satisfying $0 \le g(q) \le 1$. Note that a powerful serve corresponds to a low value of q (the serve is likely to fault) and a high value of $g(q)$ (such a serve is likely to be unreturnable, provided there is no fault), while a softer serve is unlikely to fault (q is high) but more likely to be returned if there is no fault ($g(q)$ is low). The idea is

$$P(\text{server wins}) = q\left[g(q) + (1 - g(q))P(\text{server wins rally})\right]$$

We now add a model of the rally that will ensue if the receiver manages to return the serve (over the net). During the rally, whenever the ball crosses the net, there are three possibilities: either the ball lands out of bounds, or the responding player (next hitter) cannot return the ball over the net, or the responding player returns the ball over the net and the rally continues for at least one more hit. We assume that these events occur with probabilities:

$s =$ Probability that the ball lands out of bounds,
$t =$ Probability that the ball is in-bounds but the next hitter cannot return it,
$1 - s - t =$ Probability that the next hitter hits the ball over the net.

Thus, at any stage, the probability that the rally continues with one more passage over the net is $1 - s - t$. Note that we model both players as characterized by the same probabilities so that, if their probabilities of winning the rally are in fact different, it is because of the advantage (or disadvantage) of being the player who is about to continue the rally (or not). In other words, we model the players as equally skilled in a rally, with any asymmetry arising only as a result of order—who hits first. Note that we require that $t \geq 0$, $s \geq 0$, and $t + s \leq 1$.

The first player in the rally wins if (1) she wins immediately because the second hitter cannot return the ball or (2) she does not win immediately but her ball lands in bounds and is returned by the second hitter, whose ball crosses the net but lands out of bounds or (3) both the first and second hitters hit the ball over the net once, and then the first hitter wins after the third hit of the rally because the second hitter cannot return the ball, etc. Thus, we obtain the following infinite geometric series for the probability that the first player in the rally wins:

$$h \equiv t + t(1-s-t)^2 + t(1-s-t)^4 + \dots$$
$$+ s(1-s-t) + s(1-s-t)^3 + \dots = \frac{1-s}{2-s-t}.$$

Similarly, the player who does not start the rally—the receiver in our model—wins the point with probability $1 - h = \frac{1-t}{2-s-t}$. Note that, as $t \to 1$ (which necessarily implies $s \to 0$), the first player becomes more and more certain to win the rally. On the other hand, as $s \to 1$ (which implies $t \to 0$), (s)he becomes more and more certain to lose the rally. Reducing the values of both s and t increases the expected length of the rally, which equals $\frac{1}{s+t}$. We henceforth assume that $s > 0$ and $t > 0$, which implies $0 < h < 1$. Note that $s < t$ if and only if $h < \frac{1}{2}$. If the players' rallying abilities are unequal (denote them by (t_1, s_1), (t_2, s_2)), then the first player's win probability becomes $\frac{t_1+s_2(1-s_1-t_1)}{t_1+s_1+(t_2+s_2)(1-s_1-t_1)}$.

If the server chooses a nonfault probability x, then, given that the serve does not fault, the server wins the point with conditional probability

$$f(x) = g(x) + (1 - g(x))\,h = h + (1-h)g(x).$$

This is so because, if the serve is in play (i.e., does not land out of bounds), the server wins the point if and only if either the serve is not returned or if it is returned and the server wins the subsequent rally. Exactly as for the basic model, the unconditional probability that the server wins the point is $W(x) = xf(x)$.

The problem of finding the optimal second-serve then leads to the analogue of (1), which is

$$h + (1-h)g(p) + (1-h)pg'(p) = 0,$$

which we write as

$$g(p) + pg'(p) = -\frac{h}{1-h} \equiv -r. \tag{3}$$

The dependence of p^* on t and s is now easy to see. Start with any values of t and s that correspond to $p^* < 1$. Then increasing t and/or decreasing s makes r larger, and therefore $-r$ smaller. Since the left side of (3) is a decreasing function of p, increasing t and/or decreasing s causes p^* to increase, so p^* is increasing in r. If $r > 0$ starts near 0 and increases, then $-r$ starts just below zero and decreases. If $p^* < 1$, then p increases as $-r$ decreases. As soon as $-r = g(1) + g'(1)$, and for all smaller values of $-r$, $p^* = 1$. Thus, it follows that servers who do better in rallies, with higher values of r (comprised of higher values of t and lower values of s), should choose higher and higher nonfault probabilities on their second serves. This conclusion seems intuitive since, if reaching the rally stage makes a server better off, this server should increase the probability that there is a rally, $p(1-g(p))$, which is an increasing function of the nonfault probability p.

Assuming that the second-serve strategy, if needed, will be p^*, the server's first-serve problem is to select $q \in [0, 1]$ in order to maximize $V(q) = q\,[g(q) + (1-g(q))\, p^* f(p^*)]$. Again, we solved this problem in general as part of our analysis of the basic model. The analogue of (2) is the equation solved by the unique maximizer q^* if it satisfies $q^* < 1$. This equation is

$$h + (1-h)g(q) + (1-h)qg'(q) = p^*\left[h + (1-h)g(p^*)\right],$$

which we write in the form

$$g(q) + qg'(q) = p^*g(p^*) - r(1-p^*). \tag{4}$$

In our analysis of the basic model, we showed that (4) has a solution $q^* < 1$ if and only if $f(1) + f'(1) < p^*f(p^*) - r(1-p^*)$, which is equivalent to $g(1) + g'(1) < p^*g(p^*) - r(1-p^*)$. In this case, $0 < q^* < 1$. Otherwise, if $g(1) + g'(1) \geq p^*g(p^*) - r(1-p^*)$, the optimal first-serve strategy is $q^* = 1$.

First observe that, if $p^* = 1$, then (4) becomes $g(q) + qg'(q) = g(1)$, so $q^* = 1$ only if $g'(1) = 0$, which is impossible under our assumptions. Thus, $p^* = 1$ implies $q^* < 1$. At the other extreme, as $p^* \to 0$, (4) approaches (3), so $q^* \to p^*$. Thus, if the optimal second-serve strategy has a high probability of faulting—an indication of desperation on the part of the server—optimizing the first-stage strategy means trying to achieve an immediate win.

Players who do better in rallies should choose less risky serves—both first and second.

Example. Power No-Return Probability

Let the no-return probability equal $g(x) = a - bx^c$, where a, b, and c are positive numbers satisfying $b \leq a \leq 1$ and $c \geq 1$, and note that $g(\cdot)$ is concave. With respect to the optimal second serve, note first that $p^* = 1$ if and only if $g(1) + g'(1) = a - b(1+c) \geq -r$. Therefore, if $a \geq b(1+c)$, then $p^* = 1$ for all values of r, whereas, if $a < b(1+c)$, then $p^* = 1$ if and only if $r \geq b(1+c) - a > 0$.

If $g(1) + g'(1) = a - b(1+c) < -r$, (3) applies, and $p^* \in (0, 1)$ is the unique root of $a - b(1+c)p^c = -r$. It follows that

$$p^* = \left(\frac{a+r}{b(1+c)}\right)^{\frac{1}{c}}. \tag{5}$$

It is immediate that, in this case, the interior optimum p^* is increasing in a and decreasing in b and c. Of course, it is increasing in r.

As for the first serve, suppose that $a - b(1+c) \geq -r$ so that $p^* = 1$. Then, as noted above, q^* is the unique solution of $g(q) + qg'(q) = g(1)$ or $a - b(1+c)q^c = a - b$. It follows that

$$q^* = \left(\frac{1}{1+c}\right)^{\frac{1}{c}}$$

Thus, p^* and q^* are independent of r if either $a - b(1+c) \geq 0$ or if $a - b(1+c) < 0$ and $r \geq b(1+c) - a$. Note that, in both of these cases, $0 < q^* < p^* = 1$, consistent with the Proposition.

Now suppose that $a - b(1+c) < -r$ so that p^* is given by (5). Then q^* is the unique root of (4), which because $p^* < 1$ becomes

$$a - b(1+c)q^c = -r + \frac{c(a+r)}{1+c}\left(\frac{a+r}{b(1+c)}\right)^{\frac{1}{c}}.$$

It follows that

$$q^* = \left[\frac{a+r}{b(1+c)} - \frac{c}{1+c}\left(\frac{a+r}{b(1+c)}\right)^{1+\frac{1}{c}}\right]^{\frac{1}{c}},$$

which can be written

$$q^* = p^*\left(1 - \frac{c}{1+c}p^*\right)^{\frac{1}{c}}. \tag{6}$$

Recall also, from the Proposition, that $q^* < p^*$. Both p^* and q^* are increasing in r, so they are also increasing in t and decreasing in s.

For example, suppose that $a = b = c = 1$. Then $a - b(1+c) = -1$, so the expressions for p^* and q^* depend on whether $r \geq 1$ or $0 < r < 1$, when Equations (5) and (6) apply. We find that the optimal second and first serves, respectively, are characterized by

$$(p^*, q^*) = \begin{cases} \left(\frac{1+r}{2}, \frac{3+2r-r^2}{8}\right) & \text{if } 0 < r < 1, \\ \left(1, \frac{1}{2}\right) & \text{if } r \geq 1. \end{cases}$$

Note that both q^* and p^* are continuous at $r = 1$. Our measure of the difference between q^* and p^* is the proportionate increase in fault probability of the first serve relative to the second, or $\frac{p^*-q^*}{p^*}$. This ratio approaches $\frac{1}{4} = 0.25$ as $r \longrightarrow 0$; it increases as r increases until it reaches $\frac{1}{2} = 0.50$ at $r = 1$ and remains at this value for all larger values of r.

In terms of the original rally parameters t (immediate win probability) and s (immediate loss probability), we would write p^* and q^* as

$$(p^*, q^*) = \begin{cases} \left(\frac{2-s-t}{2(1-t)}, \frac{4-8t+2st+3t^2-s^2}{8(1-t)^2}\right) & \text{if } 0 < t < s < 1, \\ \left(1, \frac{1}{2}\right) & \text{if } 0 < s \leq t. \end{cases}$$

In particular, the lowest possible value of the fault probability reduction ratio, 0.25, is approached as $s \to 1$, and the greatest possible value, 0.5, is attained whenever $s \leq t$. In summary, a server who does better in rallies takes smaller risks on both first and second serves and also makes greater reductions in risk.

Conclusions

We began by reviewing some arguments that second serves should be slower than first serves, as they are for most tournament players. Is this a good idea? Our models imply that it is. Gale's theorem (the Proposition) shows that the first serve should always have strictly higher fault probability than the second, with the only exception being in the case when both serves are maximally slow—they are put into play with certainty. The difference we found between q^* (first serve) and p^* (second serve) also answers a "what-if?" question: What would serves be like if the second-serve rule were eliminated from tennis? There would be a difference inasmuch as there is a gap in fault probability between the first and second serves (because, in the hypothetical case, the latter would be that only serve). How wide a gap? Our models do not provide us with any general bound on this difference. But the explicit numerical models we have studied suggest that it may be quite large; we found that the fault probability of the optimal first serve was typically 30% – 50% higher than the optimal second serve. Thus, empirically observed tournament strategies may not be all that far from optimal.

Of course, there are in practice many other considerations and in theory many other approaches. The serving ability of a particular player, as well as his/her rallying ability, are clearly relevant to the choice of serving strategy. Some players can deliver fast tricky serves with a high degree of accuracy, while others lack this capability. We have been able to provide some general insights into how rallying ability might affect serving strategy, but we have not addressed the effect of individual serving ability. All of these ideas might be studied in future work.

In tennis, the second chance rule applies only when the first serve faults and not when the server loses the point in a rally following an in-play first serve. Thus, the problem of selecting a serving strategy in tennis is different from two-trials problems in other settings, such as whether to take a "mulligan" in golf. Moreover, there are contexts, such as failure to meet contractual commitments, where second chances are allowed—but only after penalties have been paid. For tennis serves that fault, however, the second chance is free.

We close by noting another issue: Our analysis has been entirely decision-theoretic, but the problem of serving strategy could well have a strategic aspect for which decision theory is not entirely appropriate. A receiver who anticipates a weaker second serve would optimally move a little closer to the net. Then, assuming the serve is as expected, the receiver's return probability is increased, reducing the server's contingent win probability. Of course, if the server unexpectedly uses a strong serve, this positioning makes it harder for the receiver to respond. Variations in return strategies are often observed in tournament play, especially in women's tennis. This interactive decision problem must be studied using game-theoretic ideas, another project we leave for future research.

REFERENCES

1. Anonymous (2010). My take on second serves. http://cheaptalk.wordpress.com.AccessedSeptember1.
2. S. Baliga, Should Roger Federer's second serve be as fast as his first? (2010) http://cheaptalk.wordpress.com.AccessedAugust31.
3. J. Branch, A pointless approach: The psychology of the double fault, *New York Times*, (2010) August 29.
4. O. Cohen, *Risk Management in the Game of Tennis,* M.Sc. Project (in Hebrew), Department of Industrial Engineering, Tel-Aviv University, 2008.
5. A. B. Dixit, B. J. Nalebuff, *Thinking Strategically*. Fourth ed., W.W. Norton, 1993.
6. D. Gale, Optimal strategy for serving in tennis, *Math. Mag.* **44** no. 4 (1971) 197–199.
7. L. Gillman, Missing more serves may win more points, *Math. Mag.* **58** no. 4 (1985) 222–224.
8. J. M. Norman, Dynamic programming in tennis—When to use a fast serve, *J. Oper. Res. Soc.* **36** (1985) 75–77.
9. G. Pollard, D. Meyer, Overview of operations research in tennis, in *Wiley Encyclopedia of Operations Research and Management Science*, Wiley, 2011.

Summary. In tennis, the server has an advantage—the opportunity to serve again without penalty, if the attempt results in a fault. A common strategy is to hit a powerful or "tricky" first serve followed, if necessary, by a weaker second serve that has a lower probability of faulting even if it is easier to return. Recently, commentators have argued that this standard strategy is flawed and that the second serve should be as difficult to return as the first. This advice contradicts Gale's theorem, which we reformulate and provide with a new (analytic) proof. Then

we extend it with a model of the rally that follows the successful return of a serve, providing additional insight into the relative effectiveness of the "two first serves" strategy. The only tools we use are basic probability and introductory calculus.

YIGAL GERCHAK (MR Author ID: 72625) is a Professor at the Department of Industrial Engineering of Tel-Aviv University, Israel. Prior to that he spent many years at the University of Waterloo, Canada. His current main research interests are supply chain management, inventory pooling, principal agent models and contests. He likes to watch sports and occasionally write about them (it is less painful to the muscles than actually engaging in them).

MARC KILGOUR (MR Author ID: 101405) is Professor of Mathematics at Wilfrid Laurier University in Waterloo, Canada. His research interests include game theory, group decision and negotiation, fair division, and electoral systems. He is a fan of some sports, but he appreciates them all because they are great sources of problems, not only problems of strategy and management, but also of design—how to write rules that induce desired behavior.

Real Mathematics Facts

Snapple includes "Real Facts" on the underside of the caps to their beverages. A number of them are related to mathematics, including the following two.

Snapple's "Real Fact" #65 states that "A one-day weather forecast requires about 10 billion math calculations."

Snapple's "Real Fact" #968 states that "-40 degrees Fahrenheit is the same temperatures as -40 degrees Celsius." The linear relationship between temperature in Fahrenheit (F) and in Celsius (C) is $F = \frac{9}{5}C + 32$. Letting $F = C$, then $C = \frac{9}{5}C + 32$, or $C = -40$, verifying the fact.

More Snapple facts can be found at the website:

https://www.snapple.com/real-facts

The M&M Game: From Morsels to Modern Mathematics

IVAN BADINSKI
CHRISTOPHER HUFFAKER
NATHAN McCUE
CAMERON N. MILLER
KAYLA S. MILLER
STEVEN J. MILLER
MICHAEL STONE
Williams College
Williamstown, MA 01267
sjm1@williams.edu

The M&M Game began as a simple question asked by the sixth author's curious four-year-old son, Cam: If two people are born on the same day, do they die on the same day? Of course, needing a way to explain randomness to children (two-year-old daughter Kayla was there as well), the three took the most logical next step and used M&M'S and created the M&M Game (see Figure 1 for an illustration).

You and some friends start with some number of M&M'S. Everyone flips a fair coin at the same time; if you get a head, you eat an M&M; if you get a tail, you don't. You continue tossing coins together until no one has any M&M'S left, and whoever is the last person with an M&M lives longest and "wins"; if all run out simultaneously, the game is a tie.

Figure 1 The first M&M Game; for young players, there is an additional complication in that it matters which colors you have and the order you place them down.

We can reformulate Cam's question on randomness as follows. If everyone starts with the same number of M&M'S, what is the chance everyone eats their last M&M at the same time? We'll concentrate on two people playing with c (for Cam) and k (for Kayla) M&M'S, though we encourage you to extend to the case of more people

Math. Mag. **90** (2017) 197–207. doi:10.4169/math.mag.90.3.197. © Mathematical Association of America
MSC: Primary 65Q30, Secondary 60C05; 33C05

playing, possibly with a biased coin. In the course of our investigations, we'll see some nice results in combinatorics and see applications of memoryless processes, statistical inference, and hypergeometric functions.

Recalling that the binomial coefficient $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ denotes the number of ways to choose r objects from n when order doesn't matter, we can compute the probability $P(k, k)$ of a tie when two people start with k M&M'S. If we let $P_n(k, k)$ denote the probability that the game ends in a tie with both people starting with k M&M'S after *exactly* n moves, then

$$P(k, k) = \sum_{n=k}^{\infty} P_n(k, k).$$

Note that we are starting the sum at k as it is impossible all the M&M'S are eaten in fewer than k moves.

We claim that

$$P_n(k, k) = \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n.$$

This formula follows from the following observation: If the game ends in a tie after n tosses, then each person has *exactly* $k-1$ heads in their first $n-1$ tosses. As we have a fair coin, each string of heads and tails of length n for a player has probability $(1/2)^n$. The number of strings for each person where the first $n-1$ tosses have *exactly* $k-1$ heads, and the n^{th} toss is a head (we need this as otherwise we do not have each person eating their final M&M on the n^{th} move) is $\binom{n-1}{k-1}\binom{1}{1}$. The $\binom{1}{1}$ reflects the fact that the last toss must be a head. As there are two players, the probability that each has their k^{th} head after the n^{th} toss is the product, proving the formula. We have thus shown the following.

Theorem 1. *The probability the M&M Game ends in a tie with two people using fair coins and starting with k M&M'S is*

$$P(k, k) = \sum_{n=k}^{\infty} \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n = \sum_{n=k}^{\infty} \binom{n-1}{k-1}^2 \frac{1}{2^{2n}}. \quad (1)$$

While the above solves the problem, it is unenlightening and difficult to work with. The first difficulty is that it involves an infinite sum over n. (In general, we need to be careful and make sure any infinite sum converges; while we are safe here as we are summing probabilities, we can elementarily prove convergence. Note $\binom{n-1}{k-1} \leq n^{k-1}/(k-1)!$, and thus the sum is bounded by $(k-1)!^{-2} \sum_{n \geq k} n^{2k-2}/2^{2n}$; as the polynomial n^{2k-2} grows significantly slower than the exponential factor 2^{2n}, the sum rapidly converges.) Second, it is very hard to sniff out the k-dependence: If we double k, what does that do to the probability of a tie? It is desirable to have exact, closed-form solutions so we can not only quickly compute the answer for given values of the parameter but also get a sense of how the answer changes as we vary those inputs. In the sections below, we'll look at many different approaches to this problem, most of them trying to convert the infinite sum to a more tractable finite problem.

The basketball problem, memoryless games, and the geometric series formula

A basketball game We can convert the infinite M&M Game sum, Equation (1), into a finite sum as we have a memoryless game: The behavior of the system only depends on its state at a given moment in time and not on how we got there. There are many examples where all that matters is the configuration, not the path taken to reach it. For example, imagine a baseball game. If the lead-off hitter singles or walks, the net effect is to have a runner on first, and the two results are (essentially) the same. For another example, consider a game of Tic-Tac-Toe; what matters is where the X's and O's are on the board, not the order they are placed.

We first look at a related problem that's simpler but illustrates the same point and yields the famous geometric series formula. Imagine two of the greatest basketball players of all time, Larry Bird of the Boston Celtics and Michael Jordan of the Chicago Bulls, are playing a basketball game. The rules below are a slight modification of a famous Superbowl ad, "The Showdown," between the two where the first one to miss gives up his claim to a bag of McDonald's food to the other. In that commercial, the two keep taking harder and harder shots; for our version, we'll have them do the same shot each time. Explicitly, the rules of their one-on-one game of hoops are as follows.

> *Bird and Jordan alternate shooting free throws, with Bird going first, and the first player to make a basket wins. Assume Bird always makes a basket with probability p_B, while Jordan always gets a basket with probability p_J. If the probability Bird wins is x_B, what is x_B?*

Note that this is almost a simplified M&M Game: There is only one M&M, but the players take turns flipping their coins. We'll see, however, that it is straightforward to modify the solution and solve our original problem.

Solution from the geometric series formula The standard way to solve our basketball problem uses a geometric series. The probability that Bird wins is the sum of the probabilities that he wins on his n^{th} shot. We'll see in the analysis below that it's algebraically convenient to define $r := (1 - p_B)(1 - p_J)$, which is the probability they both miss. Let's go through the cases. We assume that p_B and p_J are not both zero; if they were, then neither can hit a basket. Not only would this mean that our ranking of them as two of the all-time greats is wrong, but the game will never end, and thus there's no need to do any analysis!

1. Bird wins on his 1^{st} shot with probability p_B.
2. Bird wins on his 2^{nd} shot with probability $(1 - p_B)(1 - p_J)p_B = rp_B$.
3. Bird wins on his n^{th} shot with probability $(1 - p_B)(1 - p_J) \cdot (1 - p_B)(1 - p_J) \cdots (1 - p_B)(1 - p_J)p_B = r^{n-1}p_B$.

To see this, if we want Bird to win on shot n, then we need to have him and Jordan miss their first $n - 1$ shots, which happens with probability $((1 - p_B)(1 - p_J))^{n-1} = r^{n-1}$, and then Bird hits his n^{th} shot, which happens with probability p_B. Thus,

$$\text{Prob(Bird wins)} \;=\; x_B \;=\; p_B + rp_B + r^2 p_B + r^3 p_B + \cdots \;=\; p_B \sum_{n=0}^{\infty} r^n,$$

which is a geometric series. As we assumed p_B and p_J are not both zero, $r = (1 - p_B)(1 - p_J)$ satisfies $|r| < 1$, and we can use the geometric series formula to deduce

$$x_B \;=\; \frac{p_B}{1-r} \;=\; \frac{p_B}{1-(1-p_B)(1-p_J)}.$$

We have made enormous progress. We converted our infinite series into a ***closed-form expression***, and we can easily see how the probability of Bird winning changes as we change p_B and p_J.

Solution through memoryless game and the geometric series formula We now give a second solution to the basketball game; instead of requiring the geometric series formula as an input, we obtain it as a consequence of our arguments.

Recall the assumptions we made. The probability Bird makes a shot is p_B, the probability Jordan hits a basket is p_J, and the probability they both miss is $r := (1 - p_B)(1 - p_J)$. We can use this to compute x_B, the probability Bird wins, in another way. Before, we wrote x_B as a sum over the probabilities that Bird won in n games. Now, we claim that

$$\text{Prob(Bird wins)} \;=\; x_B \;=\; p_B + r x_B.$$

To see this, note either Bird makes his first basket and wins (which happens with probability p_B) or he misses (with probability $1 - p_B$). If Bird is going to win, then Jordan must miss his first shot, and this happens with probability $1 - p_J$. Something interesting happens, however, if both Bird and Jordan miss: *We have reset our game to its initial state!* Since both have missed, it's as if we just started playing the game right now. Since both miss and Bird has the ball again, by definition, the probability Bird wins from this configuration is x_B, and thus the probability he wins is $p_B + (1 - p_B)(1 - p_J)x_B$.

Solving for x_B, the probability Bird beats Jordan is

$$x_B \;=\; \frac{p_B}{1 - r}.$$

As this must equal the infinite series expansion from the previous subsection, we deduce the geometric series formula:

$$\frac{p_B}{1-r} \;=\; p_B \sum_{n=0}^{\infty} r^n \quad \text{therefore} \quad \sum_{n=0}^{\infty} r^n \;=\; \frac{1}{1-r}.$$

Remark. We have to be a bit careful. It's important to keep track of assumptions. In our analysis, $r = (1 - p_B)(1 - p_J)$ with $0 \le p_B, p_J \le 1$, and at least one of p_B and p_J is positive (if both were zero the game would never end). Thus, we have only proved the geometric series formula if $0 \le r < 1$; we encourage you to find a way to pass to all $|r| < 1$.

Let's look closely at what we've done in this subsection. The key observation was to notice that we have a memoryless game. We now show how to similarly convert the solution to the M&M Game, Equation (1), into an equivalent finite sum.

Memoryless M&M Games

Setup. Remember (Equation (1)) that we have an infinite sum for the probability of a tie with both people starting with k M&M'S:

$$P(k,k) \;=\; \sum_{n=k}^{\infty} \binom{n-1}{k-1} \left(\frac{1}{2}\right)^{n-1} \frac{1}{2} \cdot \binom{n-1}{k-1} \left(\frac{1}{2}\right)^{n-1} \frac{1}{2}.$$

It's hard to evaluate this series as we have an infinite sum and a squared binomial coefficient whose top is changing. We want to convert it to something where we have more familiarity. From the hoops game, we should be thinking about how to obtain a *finite* calculation. The trick there was to notice we had a memoryless game, and all that mattered was the game state, not how we reached it. For our problem, we'll have many tosses of the coins, but in the end, what matters is where we are, not the string of heads and tails that got us there.

Let's figure out some way to do this by letting $k = 1$. In this case, we can do the same thing we did in the hoops game and boil the problem down into cases. There are four equally likely scenarios each time we toss coins, so the probability of each event occurring is 1/4 or 25%.

1. Both players eat.
2. Cam eats an M&M, but Kayla does not.
3. Kayla eats an M&M, but Cam does not.
4. Neither eat.

These four possibilities lead to the infinite series in Equation (1), as we calculate the probability the game ends in n tosses. It turns out one of the four events is not needed, and if we remove it, we can convert to a finite game.

Similar to the hoops game, we have another memoryless game. If Cam and Kayla both get tails and therefore don't eat their M&Ms, then it's as if the coin toss never happened. We can therefore ignore the fourth possibility. If you want, another way to look at this is that if we toss two tails, then there is no change in the number of M&M'S for either kid, and thus we may pretend such a toss never happened. This allows us to remove all the tosses of double tails, and now after each toss at least one player, possibly both, have fewer M&M'S. As we start with a finite number of M&M'S, the game terminates in a finite number of moves. Thus, instead of viewing our game as having four alternatives with each toss, there are only three, and they all happen with probability 1/3:

1. both players eat;
2. Cam eats an M&M, but Kayla does not;
3. Kayla eats an M&M, but Cam does not.

Notice that after each toss the number of M&M'S is decreased by either 1 or 2, so the game ends after at most $2k - 1$ tosses.

Solution We now replace the infinite sum of Equation (1) with a finite sum. Each of our three possibilities happens with probability 1/3. Since the game ends in a tie, the final toss must be double heads with both eating, and each must eat exactly $k - 1$ M&M'S in the earlier tosses. Let n denote the number of times both eat before the final toss (which again we know must be double heads); clearly, $n \in \{0, 1, \dots, k-1\}$. We thus have $n + 1$ double heads, and thus Cam and Kayla must each eat $k - (n+1) = k - n - 1$ times when the other doesn't eat.

We see that, in the case where there are $n + 1$ double heads (with the last toss being double heads), the total number of tosses is

$$(n+1) + (k-n-1) + (k-n-1) \ = \ 2k - n - 1.$$

In the first $2k - n - 2$ tosses, we must choose n to be double heads, then of the remaining $(2k - n - 2) - n = 2k - 2n - 2)$ tosses before the final toss we must choose $k - n - 1$ to be just heads for Cam, and then the remaining $k - n - 1$ tosses before the final toss must all be just heads for Kayla. These choices explain the presence of the two binomial factors. As each toss happens with probability 1/3, this explains those factors; note we could have just written $(1/3)^{2k-n-1}$, but we prefer to highlight the sources. We have thus shown the following result.

Theorem 2. *The probability the M&M Game ends in a tie with two people using fair coins and starting with k M&M'S is*

$$P(k,k) \;=\; \sum_{n=0}^{k-1} \binom{2k-n-2}{n} \left(\frac{1}{3}\right)^{n} \binom{2k-2n-2}{k-n-1} \left(\frac{1}{3}\right)^{k-n-1} \left(\frac{1}{3}\right)^{k-n-1} \frac{1}{3}. \quad (2)$$

Viewing data

Plotting Before turning to additional ways to solve the problem, it's worthwhile to pause for a bit and discuss how to view data and use results for small expressions involving k; this finite sum is certainly easier to use than the infinite sum in Equation (1), and we plot it in Figure 2 (left).

While Equation (2) gives us a nice formula for finite computations, it's hard to see the k dependence. An important skill to learn is how to view data. Frequently, rather than plotting the data as given, it's better to do a log–log plot. What this means is that, instead of plotting the probability of a tie as a function of k, we plot the logarithm of the probability of a tie against the logarithm of k. We do this in Figure 2 (right).

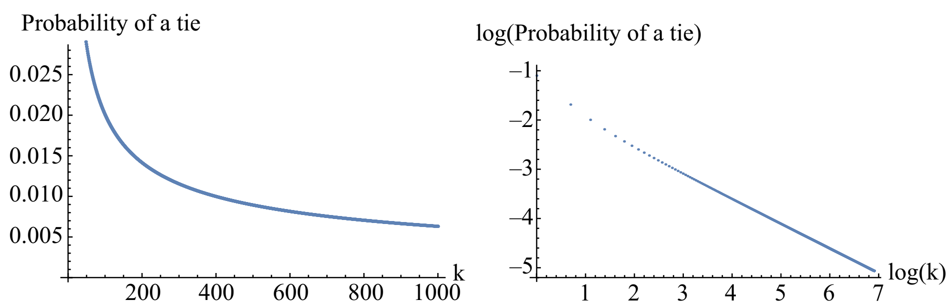


Figure 2 Left: the probability of a tie for $k \leq 1000$. Right: log–log plot.

Notice that the plot here looks *very* linear. Lines are probably the easiest functions to extrapolate, and if this linear relationship holds, we should be able to come up with a very good prediction for the logarithm of the probability (and hence, by exponentiating, obtain the probability). We do this in the next subsection.

Statistical inference Let's predict the answer for large values of k from smaller ones. The sixth named author gave a talk on this at the 110^{th} meeting of the Association of Teachers of Mathematics in Massachusetts in March 2013, which explains the prevalence of 110 and 220 below.

Figure 3 (left) gives the log-log plot for $k \leq 110$, while the right is for $50 \leq k \leq 110$. Using the Method of Least Squares with $P(k)$ the probability of a tie when we start with k M&M'S, we find a predicted best fit line of

$$\log\left(P(k)\right) \;\approx\; -1.42022 - 0.545568 \log k, \quad \text{or} \quad P(k) \;\approx\; 0.2412/k^{.5456}.$$

This predicts a probability of a tie when $k = 220$ of about 0.01274, but the answer is approximately 0.0137. While we are close, we are off by a significant amount. (In situations like this, it is better to look at not the difference in probabilities, which is small, but the percentage we are off; here, we differ by about 10%.)

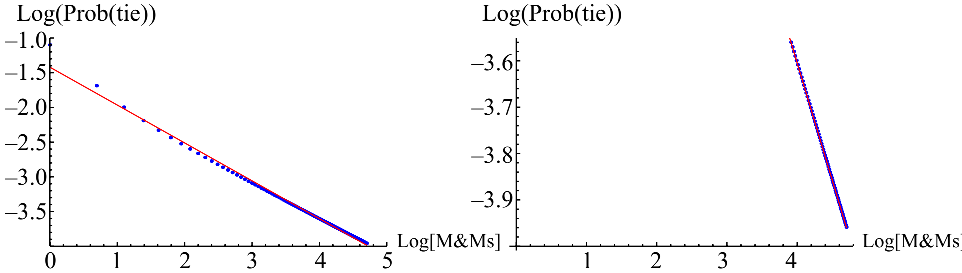


Figure 3 The probability of a tie. Left: $k \leq 110$. Right: $50 \leq k \leq 110$.

Why are we so far off? The reason is that small values of k are affecting our prediction more than they should. If we have a main term in the log–log plot that is linear, it will eventually dominate lower order terms, *but* those lower order terms could have a sizable effect for low k. Thus, it's a good idea to ignore the smaller values when extrapolating our best fit line; in Figure 3 (right), we now go from $k = 50$ to 110. Our new best fit line is

$$\log(P(k)) \approx -1.58261 - 0.50553 \log k, \quad \text{or} \quad P(k) \approx 0.205437/k^{.50553}.$$

Using this formula, we predict 0.01344 for $k = 220$, which compares *very* favorably to the true answer of 0.01347.

Recurrences, hypergeometric functions, and the OEIS

The M&M recurrence Even though we have a finite sum for the probability of a tie (Equation 2), finding that formula required some knowledge of combinatorics and binomial coefficients. We give an alternate approach that avoids these ideas. Our first approach assumes we're still clever enough to notice that we have a memoryless game, and then we remark afterward how we would have found the same formula even if we didn't realize this.

We need to consider a more general problem. We always denote the number of M&M'S Cam has with c, and Kayla with k; we frequently denote this state by (c, k). Then we can rewrite the three equally likely scenarios, each with probability 1/3, as follows:

- $(c, k) \longrightarrow (c-1, k-1)$ (double heads and both eat),
- $(c, k) \longrightarrow (c-1, k)$ (Cam gets a head, and Kayla a tail),
- $(c, k) \longrightarrow (c, k-1)$ (Cam gets a tail, and Kayla a head).

If we let $x_{c,k}$ denote the probability the game ends in a tie when we start with Cam having c M&M'S and Kayla having k, we can use the above to set up a recurrence relation (see [**3**] for a brief introduction to recurrence relations). How so? Effectively, on each turn, we move from (c, k) in exactly one of the three ways enumerated above. Now, we can use simpler game states to figure out the probability of a tie when we start with more M&Ms, as in each of the three cases we have reduced the total number of M&M'S by at least one. We thus find that the recurrence relation satisfied by $\{x_{c,k}\}$ is

$$x_{c,k} = \frac{1}{3}x_{c-1,k-1} + \frac{1}{3}x_{c-1,k} + \frac{1}{3}x_{c,k-1} = \frac{x_{c-1,k-1} + x_{c-1,k} + x_{c,k-1}}{3}. \tag{3}$$

This cannot be the full story—we need to specify initial conditions. A little thought says $x_{0,0}$ must be 1 (if they both have no M&M'S then it must be a tie), while $x_{c,0} = 0$ if $c > 0$, and similarly, $x_{0,k} = 0$ if $k > 0$ (as in these cases, exactly one of them has an M&M, and thus the game cannot end in a tie).

We have made tremendous progress. We use these initial values and the recurrence relation (3) to determine $x_{c,k}$. Unfortunately, we cannot get a simple closed form expression, but we can easily compute the values by recursion. A good approach is to compute all $x_{c,k}$ where $c + k$ equals sum fixed sum s. We've already done the cases $s = 0$ and $s = 1$, finding $x_{0,0} = 1$, $x_{0,1} = x_{1,0} = 0$.

We now move to $s = 2$. We need only find $x_{1,1}$, as we know $x_{2,0} = x_{0,2} = 0$. Using the recurrence relation, we find

$$x_{1,1} = \frac{x_{0,0} + x_{0,1} + x_{1,0}}{3} = \frac{1+0+0}{3} = \frac{1}{3}.$$

Next is the case when the indices sum to 3. Of course, $x_{0,3} = x_{3,0} = 0$, so all we need are $x_{1,2}$ and $x_{2,1}$ (which by symmetry are the same). We find

$$x_{2,1} = x_{1,2} = \frac{x_{1,1} + x_{2,0} + x_{0,2}}{3} = \frac{1/3+0+0}{3} = \frac{1}{9}.$$

We can continue to $s = 4$, and after some algebra easily obtain

$$x_{2,2} = \frac{x_{1,1} + x_{2,1} + x_{1,2}}{3} = \frac{5}{27}.$$

If we continued on with these calculations, we would find that $x_{3,3} = \frac{11}{81}$, $x_{4,4} = \frac{245}{2187}$, $x_{5,5} = \frac{1921}{19863}$, $x_{6,6} = \frac{575}{6561}$, $x_{7,7} = \frac{42635}{531441}$, and $x_{8,8} = \frac{355975}{4782969}$. The beauty of this recursion process is that we have a sure-fire way to figure out the probability of a tie at different states of the M&M game. We leave it as an exercise to the interested reader to compare the computational difficulty of finding $x_{100,100}$ by the recurrence relation versus by the finite sum of Equation (2).

We end with one final comment on this approach. We can recast this problem as one in counting weighted paths on a graph. We count the number of paths from (0, 0) to (n, n) where a path with m steps is weighted by $(1/3)^m$, and the permissible steps are (1, 0), (0, 1), and (1, 1). In Figure 4, we start with $(c, k) = (4, 4)$ and look at all the possible paths that end in (0, 0).

Remark. If we hadn't noticed it was a memoryless game, we would have found

$$x_{c,k} = \frac{1}{4}x_{c-1,k-1} + \frac{1}{4}x_{c-1,k} + \frac{1}{4}x_{c,k-1} + \frac{1}{4}x_{c,k}.$$

Straightforward algebra returns us to our old recurrence, equation (3):

$$x_{c,k} = \frac{1}{3}x_{c-1,k-1} + \frac{1}{3}x_{c-1,k} + \frac{1}{3}x_{c,k-1}.$$

This means if we did not notice initially that there was a memoryless process, doing the algebra suggests there is one!

Hypergeometric functions We end our tour of solution approaches with a method that actually prefers the infinite sum to the finite one, hypergeometric functions (see, for example, [**1, 2**]). These functions arise as the solution of a particular linear second order differential equation:

$$x(1-x)y''(x) + [c - (1-a+b)x]y'(x) - aby(x) = 0.$$

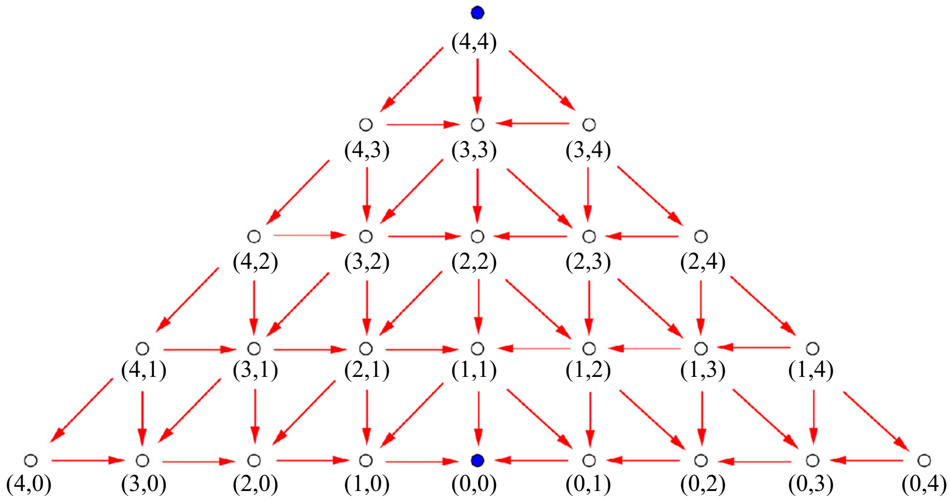


Figure 4 The M&M game when $k = 4$.

This equation is useful because every other linear second order differential equation with three singular points (in the case they are at 0, 1, and ∞) can be transformed into it. As this is a second order differential equation, there should be two solutions. One is

$$y(x) = 1 + \frac{abx}{c \cdot 1!} + \frac{a(a+1)b(b+1)x^2}{c(c+1) \cdot 2!} + \frac{a(a+1)(a+2)b(b+1)(b+2)x^3}{c(c+1)(c+2) \cdot 3!} + \cdots, \tag{4}$$

so long as c is not a nonpositive integer; we denote this solution by ${}_2F_1(a, b; c; z)$. By choosing appropriate values of a, b, and c, we recover many special functions. Wikipedia lists three nice examples:

$$\begin{aligned} \log(1+x) &= x\,{}_2F_1(1, 1; 2; -x) \\ (1-x)^{-a} &= {}_2F_1(a, 1; 1; x) \\ \arcsin(x) &= x\,{}_2F_1(1/2, 1/2; 3/2; x^2). \end{aligned} \tag{5}$$

By introducing some notation, we can write the series expansion more concisely. We define the Pochhammer symbol by

$$(a)_n = a(a+1)(a+2)\cdots(a+n-1) = \frac{(a+n-1)!}{(a-1)!}$$

(where the last equality holds for integer a; for real a, we need to interpret the factorial as its completion, the Gamma function). Our solution becomes

$${}_2F_1(a, b, c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n x^n}{(c)_n n!}.$$

Note the factorials in the above expression suggest that there should be connections between hypergeometric functions and products of binomial coefficients. In this notation, the 2 represents the number of Pochhammer symbols in the numerator, the 1 the number of Pochhammer symbols in the denominator, and the a, b, and c are what we evaluate the symbols at (the first two are the ones in the numerator, the last the denominator). One could of course consider more general functions, such as

$$ {}_sF_t(\{a_i\},\{b_j\};x) \;=\; \sum_{n=0}^{\infty} \frac{(a_1)_n\cdots(a_s)_n x^n}{(b_1)_n\cdots(b_t)_n n!}. $$

The solution ${}_2F_1(a,b,c;x)$ is called a hypergeometric function, and if you look closely at it while recalling the infinite sum solution to the M&M Game you might see the connection. After some algebra where we convert the binomial coefficients in the infinite sum solution of Equation (1) to the falling factorials that are the Pochhammer symbols, we find the following closed form solution.

Theorem 3. *The probability the M&M Game ends in a tie with two people using fair coins and starting with k M&M'S is*

$$ P(k,k) \;=\; {}_2F_1(k,k,1;1/4)4^{-k}. \tag{6} $$

It is not immediately clear that this is progress; after all, it looks like we've just given a fancy name to our infinite sum. Fortunately, special values of hypergeometric functions are well studied (see, for example, [**1, 2**]), and a lot is known about their behavior as a function of their parameters. We encourage the interested reader to explore the literature and discover how "useful" the above is.

OEIS If we use our finite series expansion of Equation (2) or the recurrence relation of Equation (3), we can easily calculate the probability of a tie for some small k. We give the probabilities for k up to 8 in Table 1. In addition, we also give $3^{2k-1}P(k,k)$ as multiplying by 3^{2k-1} clears the denominators and allows us to use the On-Line Encyclopedia of Integer Sequences (OEIS, http://oeis.org/).

TABLE 1: Probability of a tie.

k	$P(k,k)$	$3^{2k-1}P(k,k)$
1	1/3	1
2	5/27	5
3	11/81	33
4	245/2187	245
5	1921/19683	1921
6	575/6561	15525
7	42635/531441	127905
8	355975/4782969	1067925

Thus, to the M&M Game with two players, we can associate the integer sequence 1, 5, 33, 245, 1921, 15,525, 127,905, 1,067,925, ... We plug that into the OEIS and find that it knows that sequence: A084771 (see http://oeis.org/A084771, and note that the first comment on this sequence is that it equals the number of paths in the graph we discussed).

Conclusion and further questions

We've seen many different ways of solving the M&M Game, each leading to a different important aspect of mathematics. We leave the reader with some additional questions to pursue using the techniques from this and related articles.

How long do we expect a game to take? What would happen to the M&M problem if we increased the number of players? What if all of the players started with different numbers of M&M'S? What if the participants used biased coins?

In one of the first games ever played, starting with five M&M'S Kayla tossed five consecutive heads, losing immediately; years later, she still talks about that memorable performance. There is a lot known about the longest run of heads or tails in tosses of a fair (or biased) coin (see, for example, [**4**]). We can ask related questions here. What is the expected longest run of heads or tails by any player in a game? What is the expected longest run of tosses where all players' coins have the same outcome?

Acknowledgment This work was partially supported by NSF Grants DMS1265673 and DMS1561945. We thank Stan Wagon for suggesting and sharing his solution to the probability the first shooter wins the hoops game when both probabilities are independent uniform random variables, Adam Lott for help with images, and Ann Ewalt, Michael Jones, Elizabeth Miller, Frank Morgan,and the referees for comments on an earlier draft. The sixth named author thanks Donald Cameron for the opportunity to talk on this problem at the 110$^{\text{th}}$ meeting of the Association of Teachers of Mathematics in Massachusetts.

REFERENCES

1. M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965. Online at http://people.math.sfu.ca/~cbm/aands/.
2. I. Gradshteyn, I. Ryzhik, *Tables of Integrals, Series, and Products*. Academic Press, New York, 1965.
3. S. J. Miller, R. Takloo-Bighash, *An Invitation to Modern Number Theory*. Princeton Univ. Press, Princeton, NJ, 2006.
4. M. Schilling, The longest run of heads, *College Math. J.* **21** no. 3 (1990) 196–207.

Summary. To an adult, it's obvious that the day of someone's death is not precisely determined by the day of birth, but it's a very different story for a child. We invented what we call the *the M&M Game* to help explain randomness: Given k people, each simultaneously flips a fair coin, with each eating an M&M on a head and not eating on a tail. The process then continues until all M&M'S are consumed, and two people are deemed to die at the same time if they run out of M&M'S together. We analyze the game and highlight connections to the memoryless process, combinatorics, statistical inference, and hypergeometric functions.

STEVEN J. MILLER (MR Author ID: 735278) is an associate professor of mathematics at Williams College. His primary research interests are number theory and probability. He is also interested in applying math in many areas. This paper was written as part of an undergraduate probability course at Williams College.

IVAN BADINSKI (MR Author ID: 1187040), **CHRISTOPHER HUFFAKER** (MR Author ID: 1187041), **NATHAN McCUE** (MR Author ID: 1187042), and **MICHAEL STONE** (MR Author ID: 1112600) were students in Steven Miller's probability course and are now using their training in industry.

CAMERON N. MILLER (MR Author ID: 1187043) and **KAYLA S. MILLER** (MR Author ID: 1187045) were teaching assistants for the probability course. Both enjoy math, getting up early, and beating their father in Rubik's cube races.

Fontaine's Forgotten Method for Inexact Differential Equations

AARON E. LEANHARDT
Las Vegas, NV 89169
leanhardt@gmail.com

ADAM E. PARKER
Wittenberg University
Springfield, OH 45501
aparker@wittenberg.edu

Mathematics students are generally unaware of the long and convoluted paths that brought modern results into their polished form. This is through no fault of their own, as courses are often too packed with content to always travel down the historical rabbit hole. For example, ordinary differential equation (ODE) students rarely realize the wealth of sporadic, ad hoc, or forgotten methods that predated and motivated what appears in the textbook. However, those old approaches often still have value and may lead to solutions when the modern methods aren't completely general. This is what happens when we look deeper at inexact differential equations.

Today, almost every ODE student learns Alexis-Claude Clairaut's 1739 integrating factor method to solve such equations. Overall, his technique is easy to understand and applies to a wide range of examples. Interestingly, Alexis Fontaine submitted his own method of solving inexact differential equations in a 1738 paper—a paper which Clairaut was assigned to referee! Clairaut immediately saw a way to improve Fontaine's method and hence published the approach we use today.

In this note, we examine these two methods of Fontaine and Clairaut, starting by reviewing exact and inexact differential equations and their history. We'll explain how Clairaut's technique relates to solving a certain partial differential equation (PDE). We then revisit Fontaine's approach and see that it also requires a PDE analogous to, but more difficult than, Clairaut's. This PDE is the only obstruction to the use of his method and we outline Fontaine's own limitedly successful "méthode des indéterminées" to solve it. Finally, for a specific family of ODEs, we provide our own novel solutions to the resulting PDEs which could not be solved using indeterminates. Thus, we believe we have for the first time used Fontaine's method to solve an ODE which is not obviously solvable with Clairaut's approach.

History of the problem and solution techniques

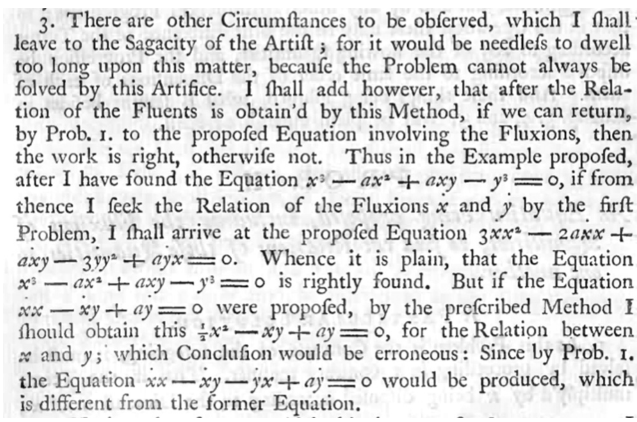
Our story begins with Newton's *Method of Fluxions and Infinite Series*, written in 1671, which describes fluxions (which evolved into today's differential calculus) and fluents (which evolved into integral calculus). The text is organized into Problems and Methods. Problem I is "The Relation of Flowing Quantities to one other being given, to determine the Relation of the Fluxions." Method I solves this by essentially computing the total differential of a relation.

More interestingly for us, Problem II is "An Equation being proposed, including the Fluxions of Quantities, to find the Relations of those Quantities to one another."

Math. Mag. **90** (2017) 208–219. doi:10.4169/math.mag.90.3.208.
MSC: Primary 34A99, Secondary 34-03; 01A50

In other words, can we solve a given differential equation of the form $Mdx + Ndy = 0$? Newton shows how to solve such an equation using series, but first describes an analytic Method II. He shows $(3x^2 - 2ax + ay)dx + (-3y^2 + ax)dy = 0$ has solution $x^3 - ax^2 + axy - y^3 = 0$.

Like any good mathematician, Newton checks his answer, using his total differentiation Method I on $x^3 - ax^2 + axy - y^3 = 0$ to recover the original differential equation. All seems fine until Newton notes his Method II does not always work! Indeed when he starts with $(x - y)dx + ady = 0$ and applies Method II he finds the solution $\frac{1}{2}x^2 - xy + ay = 0$. Applying Method I should return the original differential equation, but it does not. Why? Like any stuck mathematician, Newton leaves that to the "sagacity of the artist," the 17th century version of "left as an exercise to the reader" (see Figure 1).



3. There are other Circumſtances to be obſerved, which I ſhall leave to the Sagacity of the Artiſt; for it would be needleſs to dwell too long upon this matter, becauſe the Problem cannot always be ſolved by this Artifice. I ſhall add however, that after the Relation of the Fluents is obtain'd by this Method, if we can return, by Prob. 1. to the propoſed Equation involving the Fluxions, then the work is right, otherwiſe not. Thus in the Example propoſed, after I have found the Equation $x^3 - ax^2 + axy - y^3 = o$, if from thence I ſeek the Relation of the Fluxions $\dot{x}$ and $\dot{y}$ by the firſt Problem, I ſhall arrive at the propoſed Equation $3\dot{x}x^2 - 2a\dot{x}x + a\dot{x}y - 3\dot{y}y^2 + a\dot{y}x = o$. Whence it is plain, that the Equation $x^3 - ax^2 + axy - y^3 = o$ is rightly found. But if the Equation $\dot{x}x - \dot{x}y + a\dot{y} = o$ were propoſed, by the preſcribed Method I ſhould obtain this $\frac{1}{2}x^2 - xy + ay = o$, for the Relation between x and y; which Concluſion would be erroneous: Since by Prob. 1. the Equation $\dot{x}x - \dot{x}y - \dot{y}x + a\dot{y} = o$ would be produced, which is different from the former Equation.

Figure 1 Newton realizes his method isn't general [**14**, p. 26].

Let's be sagacious! Why does his method work in one instance and not another? It is because one of his equations is exact and the other is not, a concept not formalized until years after Newton's death. Recall that a differential equation

$$M(x, y)dx + N(x, y)dy = 0 \tag{1}$$

is *exact* if there exists a $\phi(x, y) = c$ so that

$$d\phi = \frac{\partial \phi}{\partial x}dx + \frac{\partial \phi}{\partial y}dy = Mdx + Ndy = 0.$$

Under reasonable assumptions a necessary and sufficient condition to be exact is $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, after which finding ϕ is well understood [**15**, p. 47].

Today, the canonical method of solving inexact differential equations involves finding an integrating factor $\mu(x, y)$ so that $(\mu M)dx + (\mu N)dy = 0$ is exact. Equivalently

$$\frac{\partial(\mu M)}{\partial y} = \frac{\partial(\mu N)}{\partial x},$$

which means after applying the product rule that μ solves the PDE

$$\mu\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) + M\frac{\partial \mu}{\partial y} - N\frac{\partial \mu}{\partial x} = 0. \tag{2}$$

This general technique is due to Alexis-Claude Clairaut (see Figure 2) in his 1739 "Recherches générales sur le calcul intégral" [**2**]. As is so often the case, Euler independently discovered the technique in [**5**], which was published in 1740 though written around 1734 [**11**, p. 534].

Imaginons préſentement que μ repréſente ce facteur inconnu, $\mu M dx + \mu N dy$ eſt donc la différentielle de quelque fonction de x, de y & d'une conſtante quelconque p. Donc, par notre Théoreme, la différence de μM, y variant, eſt la même que celle de μN, x variant; c'eſt-à-dire, que $\frac{d(\mu M)}{dy} = \frac{d(\mu N)}{dx}$, ou, ce qui revient au même, $\mu \frac{dM}{dy} + M \frac{d\mu}{dy} - \mu \frac{dN}{dx} - N \frac{d\mu}{dx} = 0$, Equation qui eſt d'une grande utilité pour trouver μ; car la difficulté eſt réduite à trouver la forme la plus générale que puiſſe avoir cette quantité, parce qu'à l'aide de la méthode des Indéterminées, on la déterminera à être celle qui convient pour réſoudre cette Equation.

Translation: Let us now assume that μ is this unknown. Thus $\mu M dx + \mu N dy$ is the differential of some function of x, y and some arbitrary constant p. Therefore, using our Theorem, the derivative of μM with respect to y is the same as the derivative of μN with respect to x i.e. $\frac{d(\mu M)}{dy} = \frac{d(\mu N)}{dx}$, which is the same as writing $\mu \frac{dM}{dy} + M \frac{d\mu}{dy} - \mu \frac{dN}{dx} - N \frac{d\mu}{dx} = 0$, an equation which is very useful for finding μ; since the initial difficulty has now been reduced to finding the most general form that this quantity μ can have, because with the method of undetermined coefficients, we will find μ to be what is needed to solve this Equation.

Figure 2 Clairaut gives the modern integrating factor method, followed by a translation of the original French [**2**, p. 428].

Integrating factors had been used for specific problems many times before Clairaut (or Euler) developed a general theory (we will see they even play a role in Fontaine's method). The first example is due to Fatio de Duillier in a letter to Huygens in June 1687 [**4**, p. 169] [**11**, p. 533]. He integrated $3x - 2ydx = 0$ by multiplying both sides by y^2x^{-3} to get $3y^2x^{-2} - 2y^3x^{-3}dx = 0$, which is $d(y^3x^{-2}) = 0$. Thus the solution is $y^3x^{-2} = c$. Johann Bernoulli generalized this a bit by solving $axdy - ydx = 0$ [**12**] and some other other cases. This method seems rather cumbersome for such clearly separable equations, but at the time $\int 1/xdx$ was unknown, so integrating factors circumvented this issue.

Others used integrating factors to solve specific problems, perhaps most notably Leibniz with first order linear equations [**13**, p. 257]. But these methods were all based on "intuition" [**12**, p. 211] and were not placed in a larger context until Clairaut and Euler.

Returning to our story, Clairaut's general theory was actually motivated by an earlier 1738 paper by Alexis Fontaine. Indeed Clairaut's paper [**2**] begins, "I propose to give in this paper a method to integrate differential equations, as general as that of Mr. Fontaine, but much simpler in theory and more convenient in practice" and ends with a critical "Remarques sur la méthode de M. Fontaine" which simultaneously suggests improvements while providing scant details. At that time (and also now) very few readers would be familiar with the referenced Fontaine method. In a footnote to his paper Clairaut writes, "This thesis should have appeared after that of Mr. Fontaine, titled ‘Le Calcul Integral,’ but since Mr. Fontaine did not turn his in yet, the academy decided to publish this thesis first” [**2**, p. 425]. It was not actually a question of timing as Fontaine’s paper “was not transcribed into the Royal Academy proceedings—nor was the paper ever published in the Royal Academy *Mémoires*” [**10**, p. 11]. In fact, Clairaut was only aware of it as he and François Nicole were charged with refereeing the paper, reading a negative report into the minutes in 1739 [**3**]. While his method is clearly different from Fontaine’s, Clairaut must have recognized it looked odd to reject a paper then immediately publish a solution to the same problem as he noted he could have been “accused of having disguised [Fontaine’s] ideas” [**10**, p. 34]. John L. Greenberg in [**10**] reconstructs Fontaine’s probable 1748 method from the few surviving sources.

Fontaine does publish a version of his technique almost 30 years later in the appropriately named *Mémoires Donnés à l’Académie Royal des Sciences Non Imprimés Dans Leur Temps* (*Briefs Given to the Royal Academie of Sciences, Not Printed in Their Time*) [**7**, p. 29] and reprinted verbatim in 1770 as [**8**]. Since the published versions contain (uncredited) improvements suggested by Clairaut, Greenberg states it “is practically worthless as a means for determining what Fontaine’s efforts of 1738 at integration of Equation (1) had been” [**10**, p. 33]. In actuality, Greenberg’s construction of the probable 1738 version has many similarities to the later published versions.

In the 30 years that passed between the rejection and publication of Fontaine’s method, Clairaut’s method was the only technique available. Indeed Euler used it to great effect, featuring it prominently in his “De integreatione aequationum differentialium” which was read to the St. Petersburg Academy in 1757 [**6**]. Fontaine’s method never found an audience and to our knowledge was never actually used to solve a differential equation.

Fontaine’s method

We begin by describing Fontaine’s method abstractly, but since it is rather involved we conclude this section with an illustrative example. Fontaine started with a differential equation of the form

$$dx + \alpha(x, y)dy = 0. \tag{3}$$

Clearly (3) and (1) are related by $\alpha = \frac{N}{M}$. Fontaine assumed that α was not homogeneous of degree 0 as Johann Bernoulli showed in 1728 that if M and N were both homogeneous of degree r (meaning $F(tx, ty) = t^r F(x, y)$) the change of variables $x = uy$ makes (3) separable [**1**, p. 176]. By carefully choosing a constant p in the non-homogeneous α and considering it as a variable, $\alpha(x, y, p)$ becomes homogeneous of degree 0 in x, y and p. For example $\alpha = \frac{2x^2+y^3}{4xy}$ can be made homogeneous of degree 0 by writing it as $\frac{2px^2+y^3}{4xyp}$ (and letting $p = 1$ recover the original α), or $\frac{px^2+y^3}{2xyp}$ (whereby $p = 2$ recovers α). On one hand, the problem seems more difficult in three dimensions, but he now has access to results about homogeneity, as we see below.

Returning to the method at hand, Fontaine then finds a degree 0 homogeneous function $\beta(x, y, p)$ and an integrating factor $\mu(x, y, p)$ so that

$$\mu dx + \mu\alpha dy + \mu\beta dp = 0$$

is exact, meaning there exists a function $\phi(x, y, p)$ so that

$$d\phi = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial p}dp = \mu dx + \mu\alpha dy + \mu\beta dp. \tag{4}$$

Once ϕ is known, setting p back to the original constant value with $\phi = c$ gives a solution to Equation (3). Fontaine thus solves a more general problem, and dehomogenizes the answer with respect to the variable p.

According to Greenberg in [**10**], Fontaine accomplished all this with three equations. He labeled them in the order that Clairaut and Nicole discussed them though we present them in a more logical order. The most important is analogous to Equation (2). Given α, Fontaine used it to find β. In his review, Clairaut actually complimented this useful formula, but criticized Fontaine's "si long & si difficile" proof. So he presented his own [**2**, p. 435], which follows.

Fontaine's Third Equation. If $\mu dx + \mu\alpha dy + \mu\beta dp = 0$ is exact then

$$\alpha\frac{\partial\beta}{\partial x} - \beta\frac{\partial\alpha}{\partial x} + \frac{\partial\alpha}{\partial p} - \frac{\partial\beta}{\partial y} = 0. \tag{5}$$

Proof. If $\mu dx + \mu\alpha dy + \mu\beta dp = 0$, then a necessary and sufficient condition to be exact is that $\frac{\partial\mu}{\partial y} = \frac{\partial\mu\alpha}{\partial x}$, $\frac{\partial\mu}{\partial p} = \frac{\partial\mu\beta}{\partial x}$, and $\frac{\partial\mu\alpha}{\partial p} = \frac{\partial\mu\beta}{\partial y}$, or equivalently

$$\frac{\partial\mu}{\partial y} = \mu\frac{\partial\alpha}{\partial x} + \alpha\frac{\partial\mu}{\partial x}, \qquad \frac{\partial\mu}{\partial p} = \mu\frac{\partial\beta}{\partial x} + \beta\frac{\partial\mu}{\partial x}, \qquad \mu\frac{\partial\alpha}{\partial p} + \alpha\frac{\partial\mu}{\partial p} = \mu\frac{\partial\beta}{\partial y} + \beta\frac{\partial\mu}{\partial y}.$$

Substituting the first and second into the the third gives

$$\mu\frac{\partial\alpha}{\partial p} + \alpha\left(\mu\frac{\partial\beta}{\partial x} + \beta\frac{\partial\mu}{\partial x}\right) = \mu\frac{\partial\beta}{\partial y} + \beta\left(\mu\frac{\partial\alpha}{\partial x} + \alpha\frac{\partial\mu}{\partial x}\right)$$

and cancellation gives (5). ∎

The strong homogeneity assumptions throughout Fontaine's process enabled him to utilize the well-known homogeneous function theorem, often called Euler's homogeneous function theorem. The two-dimensional theorem is indeed found in Euler [**5**, p. 185] and in Fontaine (see [**7**, p. 28] [**9**, p. 282]):

Fontaine's First Equation. $F(x, y, p)$ is a homogeneous function of degree r if and only if

$$rF = \frac{\partial F}{\partial x}x + \frac{\partial F}{\partial y}y + \frac{\partial F}{\partial p}p.$$

Since α and β are homogeneous of degree 0,

$$0 = \frac{\partial\alpha}{\partial x}x + \frac{\partial\alpha}{\partial y}y + \frac{\partial\alpha}{\partial p}p \qquad \text{and} \qquad 0 = \frac{\partial\beta}{\partial x}x + \frac{\partial\beta}{\partial y}y + \frac{\partial\beta}{\partial p}p.$$

Moreover, if ϕ is homogeneous of degree r, then

$$r\phi = \frac{\partial\phi}{\partial x}x + \frac{\partial\phi}{\partial y}y + \frac{\partial\phi}{\partial p}p = \mu x + \mu\alpha y + \mu\beta p. \tag{6}$$

This means it is trivial to find ϕ if we know α, β, and μ. In order to find μ we need Fontaine's second equation.

Fontaine's Second Equation. μ must satisfy

$$\frac{d\mu}{\mu} = \frac{(r-1-y\frac{\partial\alpha}{\partial x}-p\frac{\partial\beta}{\partial x})}{x+\alpha y+\beta p}dx + \frac{((r-1)\alpha-y\frac{\partial\alpha}{\partial y}-p\frac{\partial\beta}{\partial y})}{x+\alpha y+\beta p}dy + \frac{((r-1)\beta-y\frac{\partial\alpha}{\partial p}-p\frac{\partial\beta}{\partial p})}{x+\alpha y+\beta p}dp. \quad (7)$$

Proof. Using Equation (6) we know that

$$rd\phi = d(\mu\cdot(x+\alpha y+\beta p)) = d\mu\cdot(x+\alpha y+\beta p)+\mu\cdot d(x+\alpha y+\beta p),$$

and using Equation (4) we can eliminate $d\phi$ to get

$$\begin{aligned}\frac{d\mu}{\mu} &= \frac{r}{x+\alpha y+\beta p}dx + \frac{r\alpha}{x+\alpha y+\beta p}dy + \frac{r\beta}{x+\alpha y+\beta p}dp - \frac{d(x+\alpha y+\beta p)}{x+\alpha y+\beta p}\\ &= \frac{(r-1-y\frac{\partial\alpha}{\partial x}-p\frac{\partial\beta}{\partial x})}{x+\alpha y+\beta p}dx + \frac{((r-1)\alpha-y\frac{\partial\alpha}{\partial y}-p\frac{\partial\beta}{\partial y})}{x+\alpha y+\beta p}dy\\ &\quad + \frac{((r-1)\beta-y\frac{\partial\alpha}{\partial p}-p\frac{\partial\beta}{\partial p})}{x+\alpha y+\beta p}dp.\end{aligned}$$

■

In summary, Fontaine's third equation allows us to find β, the second allows us to find the integrating factor μ and the first allows us to find ϕ. Clairaut simplified the process by simply dividing Equation (4) by Equation (6) which Fontaine used in his 1764 version. Hence the entire process can be carried out without finding μ (though it can easily be recovered since $\partial\phi/\partial x = \mu$).

Clairaut's Improved Third Equation. [**2**, p. 434] [**7**, p. 30] ϕ must satisfy

$$\frac{1}{r}\frac{d\phi}{\phi} = \frac{1}{x+\alpha y+\beta p}dx + \frac{\alpha}{x+\alpha y+\beta p}dy + \frac{\beta}{x+\alpha y+\beta p}dp. \quad (8)$$

Luckily both Equations (7) and (8) are exact and so if we are able to find β via Equation (5) we can theoretically find ϕ by a sequence of integrations and differentiations. As noted above, most ODE textbooks derive a solution of

$$\phi = \int_x A + \int_y B - \int_y \frac{\partial}{\partial y}\int_x A = c$$

for an exact $Adx + Bdy = 0$ [**15**, p. 47] (where $\int_x$ refers to integration with respect to x etc.). The same process can be applied to the exact equation $Adx + Bdy + Cdz = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial z}dz = 0$, which gives the much less well known

$$\phi = \int_x A + \int_y B + \int_z C - \int_y \frac{\partial}{\partial y}\int_x A - \int_z \frac{\partial}{\partial z}\int_x A - \int_z \frac{\partial}{\partial z}\int_y B + \int_z \frac{\partial}{\partial z}\int_y \frac{\partial}{\partial y}\int_x A = c.$$

Let us now examine Fontaine's entire process in detail.

Example. Consider the differential equation

$$dx + \frac{2x+4}{x+y}dy = 0.$$

By considering the constant 4 as a variable p, $\alpha = \frac{2x+p}{x+y}$ becomes homogeneous of degree zero. In order to find β, Fontaine must solve Equation (5)

$$\frac{p}{x+y}\frac{\partial\beta}{\partial x} + \beta\left(\frac{p}{(x+y)^2}\right) + \left(\frac{1}{x+y}\right) - \frac{\partial\beta}{\partial y} = 0.$$

This seems hopeless, though in the next section we describe a method of Fontaine for solving this PDE in simple cases such as this. It gives $\beta = \frac{1}{2}\left(\frac{-x+y-p}{x+y}\right)$. Using the exact Equation (8) to find ϕ gives

$$\phi = ((p+2x)^{\frac{1}{3}}(x+3y-p)^{\frac{2}{3}})^r.$$

Before we check our answer, we note the role of r in Fontaine's construction. If ϕ is exact, then

$$d\phi = \frac{\partial\phi}{dx}dx + \frac{\partial\phi}{dy}dy + \frac{\partial\phi}{dp}dp = 0,$$

and any power of ϕ is also exact since

$$d(\phi^r) = r(\phi)^{r-1}d(\phi) = r(\phi)^{r-1}(0) = 0.$$

In our example it is "best" to let $r = 3$ so $\phi = (p+2x)(x+3y-p)^2 = c$. Then

$$\begin{aligned}
d\phi &= \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial p}dp \\
&= -6(p-x-3y)(x+y)dx + 6(p+2x)(-p+x+3y)dy \\
&\quad + 3(p-x-3y)(p+x-y)dp \\
&= -6(p-x-3y)(x+y)\left(dx + \frac{-6(p+2x)(p-x-3y)}{-6(p-x-3y)(x+y)}dy\right. \\
&\quad \left. + \frac{3(p-x-3y)(p+x-y)}{-6(p-x-3y)(x+y)}\right) \\
&= -6(p-x-3y)(x+y)\left(dx + \frac{(2x+p)}{(x+y)}dy + \frac{1}{2}\frac{(x-y+p)}{(x+y)}\right) \\
&= \mu\cdot(dx + \alpha dy + \beta dp).
\end{aligned}$$

Substituting $p = 4$ (which would make $dp = 0$) gives that $\phi = (2x+4)(x+3y-4)^2 = c$ is a solution to $dx + \frac{(2x+4)}{(x+y)}dy = 0$ as desired.

This problem could have been solved with our modern technique, and would have required solving Equation (2)

$$-\mu + (x+y)\frac{\partial\mu}{\partial y} - (2x+4)\frac{\partial\mu}{\partial x} = 0.$$

Historical approaches to the PDE

Fontaine knew solving Equation (5) was the key to the method and made assumptions to simplify the computation. Clairaut and Nicole state in their referee report, “The difficulty is thus reduced to determining the form of this function β. For that, Mr. Fontaine advances a proposition that the denominator of β will always be the same as that of α except that it may be multiplied by p, the numerator being the most general function of the same degree that is the denominator” [3, p. 21] (see Figure 3). They used π as our β. They continue, “. . . we recognized a large number of examples for which the theorem is not generally true.”

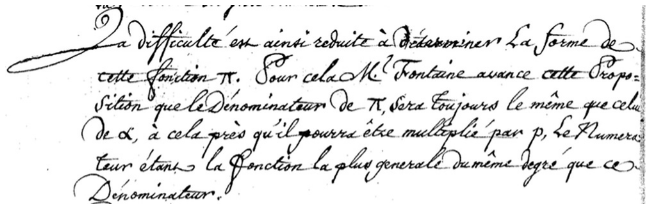

La difficulté est ainsi reduite à determiner La forme de cette fonction π. Pour cela Mr Fontaine avance cette Proposition que le Dénominateur de π, sera toujours le même que celui de α, à cela près qu'il pourra être multiplié par p, Le Numerateur étant la fonction la plus generale du même degré que ce Dénominateur.

Figure 3 Clairaut and Nicole highlight the (incorrect) assumption that Fontaine made to simplify Equation (5) [3, p. 21].

Indeed not all pairs α and β have denominators that are multiples of each other. For example if $\phi = xyp$, then $d(xyp) = ypdx + xpdy + xydp$ and so

$$\mu = yp \qquad \alpha = \frac{x}{y} \qquad \beta = \frac{x}{p}$$

and the denominator of β is no multiple of the denominator of α. However, when for given Q, $\alpha = N/M$ and $\beta = L/(QM)$, a “méthode des indéterminées” will solve the PDE. Fontaine illustrates this with two examples in his 1764 version, the first which assumes $\alpha = N/M$ and $\beta = L/M$ where M, N, and L are all homogeneous of degree 1. We substitute α and β into Equation (5) to get

$$N\frac{\partial L}{\partial x} - L\frac{\partial N}{\partial x} + M\frac{\partial N}{\partial p} - N\frac{\partial M}{\partial p} + L\frac{\partial M}{\partial y} - M\frac{\partial L}{\partial y} = 0. \tag{9}$$

We then let $N = ax + by + cp$, $M = dx + ey + fp$, and assume $L = Ax + By + Cp$ [7, p. 32]. Plugging into Equation (9) and noting the coefficients of x, y, and p must each equal zero gives a system of three equations in three unknowns

$$\begin{aligned} -Bd + cd + Ae - af &= 0, \\ Ab - aB + ce - bf &= 0, \\ Ac - aC + Ce - Bf &= 0, \end{aligned}$$

and as long as $bd - ae \neq 0$ and $a \neq e$,

$$A = \frac{a^2 f - acd - bdf + cde}{ae - bd}, \qquad B = \frac{abf - bcd - bef + ce^2}{ae - bd},$$

$$C = \frac{acf - bf^2 - c^2 d + cef}{ae - bd},$$

which completely determines the form of $\beta = L/M$ (see Figure 4). Fontaine has a typo in his coefficients and did not integrate either Equation (7) or Equation (8) to find ϕ [**7**, p. 32], yet this is as close to an explicit example using Fontaine's method that we found. Notice this method fails for $\alpha = x/y$ above. Fontaine does begin another example with $\alpha = N/M$ and $\beta = L/(pM)$ where M, N, and L are all homogeneous of degree 2. The calculations take up several pages and Fontaine ends without finding L. This "méthode des indéterminées" is the way that β was found in the above example and is also mentioned by Clairaut in Figure (2) as a means to find μ in Equation (2).

Soit, par exemple, l'équation $dx + \frac{ap+bx+cy}{\alpha p+\beta x+\gamma y}dy = 0$;
je ſuppoſe $Q = 1$; j'aurai $N = ap + bx + cy$, $M = \alpha p + \beta x + \gamma y$, $L = Ap + Bx + Cy$, $\frac{dN}{dx} = b$, $\frac{dN}{dp} = a$, $\frac{dM}{dy} = \gamma$, $\frac{dM}{dp} = \alpha$, $\frac{dL}{dx} = B$, $\frac{dL}{dy} = C$;
& en ſubſtituant dans l'équation
$N\frac{dL}{dx} - L\frac{dN}{dx} + M\frac{dN}{dp} - N\frac{dM}{dp} - M\frac{dL}{dy} + L\frac{dM}{dy} = 0$;
j'aurai $(aB - \alpha C - bA + \gamma A)\,p + (-b\alpha + \beta a - \beta C + \gamma B)\,x + (cB - c\alpha + a\gamma - bC)\,y = 0$; donc $aB - \alpha C - (b - \gamma)\,A = 0$, $-\beta C + \gamma B = b\alpha - a\beta$, $cB - bC = c\alpha - a\gamma$;
donc $A = \frac{-a(b\alpha - a\beta)}{c\beta - b\gamma}$, $B = \frac{\beta(c\alpha - a\gamma) - b(b\alpha - a\beta)}{c\beta - b\gamma}$,
$C = \frac{\gamma(c\alpha - a\gamma) - c(b\alpha - a\beta)}{c\beta - b\gamma}$;

Figure 4 Fontaine uses his "méthode des indéterminées" to find β for a simple α [**7**, p. 32].

The natural question is if Fontaine's method is ever superior to our modern Clairaut method. While Equation (2) certainly looks simpler than Equation (5), instances in which Equation (5) can be solved where neither Equation (2) nor Fontaine's "méthode des indéterminées" can be applied would be quite interesting. In the next section we use a completely different procedure to find a novel family of such examples—our family does not have denominators that are multiples of each other nor is the corresponding Equation (2) obviously solvable. We hope this paper elicits other such examples!

Our family of solutions to the PDE

Notice that the first two terms of Equation (5)

$$\alpha\frac{\partial\beta}{\partial x} - \beta\frac{\partial\alpha}{\partial x} + \frac{\partial\alpha}{\partial p} - \frac{\partial\beta}{\partial y} = 0$$

can be combined to give

$$\alpha^2\frac{\partial}{\partial x}\left(\frac{\beta}{\alpha}\right) + \frac{\partial\alpha}{\partial p} - \frac{\partial\beta}{\partial y} = 0.$$

This suggests we search for solutions of the form $\beta = \alpha^k$ and guarantees that if α is homogeneous of degree zero, then β will be as well. Substituting our guess into Equation (5) yields

$$(k-1)\alpha^k \frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial p} - k\alpha^{k-1}\frac{\partial \alpha}{\partial y} = 0. \tag{10}$$

We may be tempted to set one of the partial derivatives to zero. But if either $\frac{\partial \alpha}{\partial y} = 0$ or $\frac{\partial \alpha}{\partial x} = 0$, then the original equation was separable; and if $\frac{\partial \alpha}{\partial p} = 0$, then α was already homogeneous of degree 0 and could be separated. Nonetheless, if such solutions exist, we could form linear combinations and hope the new functions would also be solutions. This is not guaranteed because our PDE is nonlinear, but specific such solutions do behave linearly! We first set $\frac{\partial \alpha}{\partial p} = 0$ in Equation (10) which yields

$$\alpha = \frac{k}{k-1}\frac{\frac{\partial \alpha}{\partial y}}{\frac{\partial \alpha}{\partial x}}.$$

Since we are looking for homogeneous functions of degree zero, we guess that solutions are of the form $\alpha = f(\theta),\ \theta = \frac{x}{y}$. Then

$$\alpha = \frac{k}{k-1}\frac{\frac{\partial \alpha}{\partial \theta}\frac{\partial \theta}{\partial y}}{\frac{\partial \alpha}{\partial \theta}\frac{\partial \theta}{\partial x}} = -\frac{k}{k-1}\frac{x}{y}.$$

Thus, one set of solutions ($k \neq 1$) is

$$\alpha_1(k) = -\frac{k}{k-1}\frac{x}{y}, \qquad \beta_1(k) = \alpha_1^k.$$

Similarly, we set $\frac{\partial \alpha}{\partial y} = 0$ in Equation (10) and guess that solutions are of the form $\alpha = f(\theta),\ \theta = \frac{x}{p}$, which yields

$$\beta = \alpha^k = -\frac{1}{k-1}\frac{\frac{\partial \alpha}{\partial p}}{\frac{\partial \alpha}{\partial x}} = \frac{1}{k-1}\frac{x}{p}.$$

Thus, another set of solutions ($k \neq 1$) is

$$\beta_2(k) = \frac{1}{k-1}\frac{x}{p}, \qquad \alpha_2(k) = \beta_2^{\frac{1}{k}}.$$

We return to Equation (5) and search for linear combinations of our pair of solutions, $\{\alpha = \alpha_1(m) + \alpha_2(n),\ \beta = \beta_1(m) + \beta_2(n)\}$, that still satisfy the full PDE:

$$\alpha\frac{\partial \beta}{\partial x} - \beta\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial p} - \frac{\partial \beta}{\partial y} = 0.$$

Ultimately, this reduces to searching for powers $\{m,\ n\}$, that cause the "cross terms" generated by the first two terms in the PDE to cancel, i.e., we need to find $\{m,\ n\}$ such that

$$\alpha_1(m)\frac{\partial \beta_2(n)}{\partial x} - \beta_2(n)\frac{\partial \alpha_1(m)}{\partial x} + \alpha_2(n)\frac{\partial \beta_1(m)}{\partial x} - \beta_1(m)\frac{\partial \alpha_2(n)}{\partial x} = 0.$$

The four cross terms are

$$\alpha_1(m)\frac{\partial \beta_2(n)}{\partial x} = -\frac{m}{m-1}\frac{1}{n-1}\frac{x}{yp},$$

$$\beta_2(n)\frac{\partial \alpha_1(m)}{\partial x} = -\frac{m}{m-1}\frac{1}{n-1}\frac{x}{yp},$$

$$\alpha_2(n)\frac{\partial\beta_1(m)}{\partial x} = \frac{(-1)^m m^{m+1}}{(m-1)^m}\frac{1}{(n-1)^{\frac{1}{n}}}\frac{x^{m-1}x^{\frac{1}{n}}}{y^m p^{\frac{1}{n}}},$$

$$\beta_1(m)\frac{\partial\alpha_2(n)}{\partial x} = \frac{(-1)^m m^m}{(m-1)^m}\frac{\frac{1}{n}}{(n-1)^{\frac{1}{n}}}\frac{x^m x^{\frac{1}{n}-1}}{y^m p^{\frac{1}{n}}}.$$

Conveniently, the first pair of cross terms cancel for all choices of powers $\{m, n\}$, provided $m \neq 1$ and $n \neq 1$, and the remaining pair of cross terms satisfy

$$\alpha_2(n)\frac{\partial\beta_1(m)}{\partial x} - \beta_1(m)\frac{\partial\alpha_2(n)}{\partial x} = 0$$

provided $mn = 1$. The following theorem summarizes the results.

Theorem. If

$$\alpha(m) = \overbrace{\left(\frac{m}{1-m}\right)\left(\frac{x}{y}\right)}^{\alpha_1(m)} + \overbrace{\left(\frac{m}{1-m}\right)^m\left(\frac{x}{p}\right)^m}^{\alpha_2(1/m)},$$

then

$$\beta(m) = \overbrace{(-1)^m\left(\frac{m}{m-1}\right)^m\left(\frac{x}{y}\right)^m}^{\beta_1(m)} + \overbrace{\left(\frac{m}{1-m}\right)\left(\frac{x}{p}\right)}^{\beta_2(1/m)}$$

will solve Fontaine's third equation (Equation (5)) provided $m \neq 0, 1$.

As mentioned above, finding β for a given α is the main obstruction to using Fontaine's method. The remaining step in finding the solution ϕ to the original differential equation is carrying out the integrations and differentiations to solve the exact Equation (8). Thus for our α we find the explicit solution the original differential equation.

Corollary. If α and β have the above form, then a homogeneous solution of degree r to $\mu dx + \mu\alpha dy + \mu\beta dp = 0$ is given by

$$\phi = \left(\frac{(m+1)^2 yp}{x}\right)^{\frac{rm}{m+1}}\left(\frac{m^m x^m}{(1-m)^{m-1}}\left(\frac{y}{p^m} + \frac{p}{y^m}\right) + mx + x\right)^{\frac{r}{m+1}} = c.$$

With the exception of some modest progress by Fontaine himself, we know of no other instance where his method was used to analytically solve an inexact ODE. Two hundred and seventy five years is too long to have a single technique to attack these equations. We've shown there are families of ODEs which can be solved with Fontaine's method (but with no obvious way of using Clairaut's). It is our hope that this paper may motivate others to produce additional such ODEs.

Acknowledgment The authors thank Catherine Kublik and Moez Ben-Azzouz for their help with the translations and Michael Mattison for his help with the approach of the paper. In addition, they must sincerely thank the referees for their many helpful suggestions which greatly improved the paper, along with the editor who guided the manuscript through the process.

Summary. In 1739 Alexis-Claude Clairaut published the modern integrating factor method of solving inexact ordinary differential equations (ODEs). He was motivated by a 1738 Alexis Fontaine paper with a different

method which requires solving a difficult partial differential equation (PDE). Here we revisit Fontaine's method, examine his modest attempt to solve the PDE, and utilize a different technique to give (we believe) the first family of ODEs solvable by Fontaine's method with no obvious solution using the modern technique.

REFERENCES

1. J. Bernoulli, De integrationibus aequationum differentilium ubi traditur methodi alicuius specimen integrandi sine praevia separatione indeterminatarum, *Comm. Acad. Sci. Petrop.* **1** (1728) 167–184.
2. A. C. Clairaut, Recherches générales sur le calcul intégral, *Hist. Acad. R. Sci. Avec Mem. Math. Phys. Registres Cette Acad.* (1739) 425–436.
3. A. C. Clairaut, F. Nichole, Mrs. Nicole en Clairaus ons parlé ainsi sur un Ecris de M. Fontaine (sic.) Le Calcul Integral, *Acad. R. Sci. Registre Procès-Verbaux* (1739) 20–21.
4. F. deDuillier, Letter from deDuillier to Huygens, No. 2465, in *Oeuvres complètes de Christiaan Huygens.* Vol. 9 Societé Holliandaise Des Sciences, 1901.
5. L. Euler, De infinitis curvis eiusdem generis seu methodus inveniendi aequationes pro infinitis curvis eiusdem generis, *Comm. Acad. Petrop.* **7** (1740) 174–189.
6. ———, De integratione aequationum differentialium, *Novi Comm. Acad. Petrop.* **8** (1763) 3–63.
7. A. Fontaine, *Mémoires Donnés à l'Académie Royale des Sciences, Non Imprimés Dans Leur Temps.* Paris, 1764.
8. ———, *Traité de Calcul Différentiel et Intégral.* Paris, 1770.
9. J. L. Greenberg, Alex Fontaine's 'fluxio-differential method' and the origins of the calculus of several variables, *Ann. Sci.* **38** (1981) 251–290.
10. ———, Alexis Fontaine's integration of ordinary differential equations and the origins of the calculus of several variables, *Ann. Sci.* **39** (1982) 1–36.
11. E. L. Ince, *Ordinary Differential Equations.* Dover Press, New York, 1944.
12. D. Jardine, The origins of integrating factors, in *Mathematical Time Capsules. Historical Modules for the Mathematics Classoom.* Ed. D. Jardine and A. Shell-Gellasch. No. 77. MAA Notes, MAA Press, Washington DC, 2011.
13. G. W. Leibniz, Mathematische Schriften, in *Briefwechsel Zwischen Leibniz, Huygens van Zulichem, un dem Marquis de l'Hospital.* Ed. C. I. Gerhardt. Halle, 1850.
14. I. Newton, *The Method of Fluxions and Infinite Series.* John Nourse, 1736.
15. D. Zill, *A First Course in Differential Equations: The Classic Fifth Edition.* Cengage, Pacific Grove, CA, 2000.

AARON E. LEANHARDT (MR Author ID: 1190529) earned a B.S.E. in Electrical Engineering from the University of Michigan in 1999 and a Ph.D. in Physics from the Massachusetts Institute of Technology in 2003. From 2004–2007, he worked as a postdoctoral research associate at JILA, a joint institute between the National Institute of Standards and Technology and the University of Colorado. In 2007, Aaron moved to the physics faculty at the University of Michigan where he remained until 2014. Aaron is currently on "sabbati-cation," during which he has biked the Colorado Trail, broken his ankle, adopted a puppy, and become an amateur mathematician.

ADAM E. PARKER (MR Author ID: 812957) earned degrees in Mathematics and Psychology from the University of Michigan in 1999, and his Ph.D. in Algebraic Geometry from the University of Texas at Austin in 2005. Parker teaches a wide range of classes and often incorporates primary sources in his teaching. In his spare time he enjoys cooking, eating, spending time with his wife Bernadette, and repairing mechanical watches.

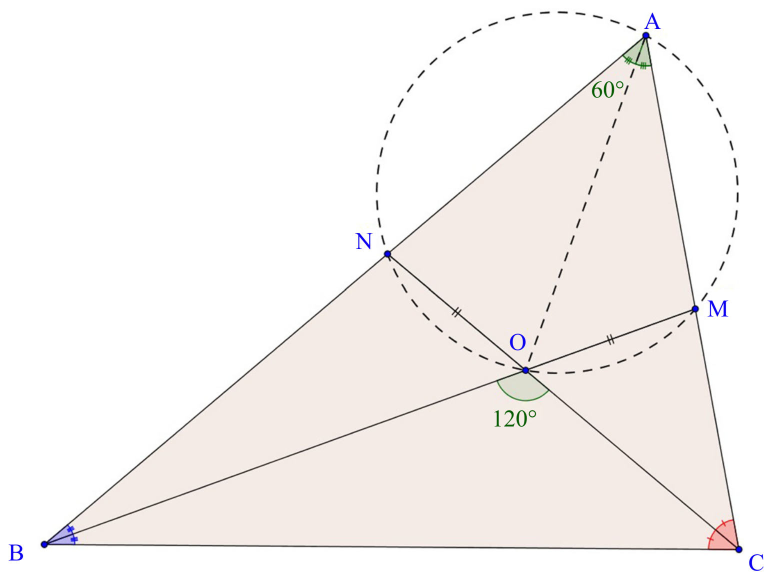
Proof Without Words: An Elegant Property of a Triangle Having an Angle of 60 Degrees

VICTOR OXMAN
Western Galilee College
Acre, 24121, Israel
victor.oxman@gmail.com

MOSHE STUPEL
Shaanan College
Haifa, 26109, Israel
stupel@bezeqint.net

For a triangle ABC with $\angle A = 60^\circ$ and with angle bisectors BM and CN that intersect at point O, then OM = ON.

Proof.



■

Exercise. Show that $OM = ON = \sqrt{\frac{1}{3}\left(a^2 + b^2 - ab\right)}$, where $a = AM$ and $b = AN$.

Summary. In a triangle ABC in which angle A measures 60 degrees, the bisectors of angles B and C are used to construct a cyclic quadrilateral with two congruent sides.

VICTOR OXMAN (MR Author ID: 756975) teaches at Western Galilee College. His research interests are in geometry and mathematical education.

MOSHE STUPEL (MR Author ID: 1028564) teaches at Shaanan College. His research interests are also in geometry and mathematical education.

Math. Mag. **90** (2017) 220. doi:10.4169/math.mag.90.3.220.
MSC: Primary 51M04

Some Probabilistic Interpretations of the Multinomial Theorem

KULDEEP KUMAR KATARIA
Indian Institute of Technology Bombay
Powai, Mumbai-400076, India
kulkat@math.iitb.ac.in

In this note, we provide a simple probabilistic proof that the number of terms in the multinomial expansion of $(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n$ is $\binom{n+m-1}{n}$. A combinatorial proof of this result using stars and bars is given in [**2**, p. 38]. By using the convolution property of the negative binomial distribution, we first obtain a more general combinatorial identity and observe that the above result follows as a special case. Moreover, we provide an alternate proof of the multinomial theorem using the convolution property of the Poisson distribution. To the best of our knowledge, these probabilistic proofs are new. For the sake of completeness and clarity, we start with the definition of the negative binomial distribution and state some results related to it.

A random variable X is said to have the negative binomial distribution with parameters $r \geq 1$ (a fixed positive integer) and $0 < p < 1$, denoted by $X \sim \mathrm{NB}(r, p)$, if its probability mass function is given by

$$\mathbb{P}\{X = k\} = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \ldots.$$

For $r = 1$, the negative binomial distribution reduces to a special case, the geometric distribution. The probability distribution of a random variable X is completely determined by its characteristic function, which is defined as the expected value of e^{itX}, $i = \sqrt{-1}$ and $t \in \mathbb{R}$. It can be shown that the characteristic function of a negative binomial random variable is $\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = p^r(1-(1-p)e^{it})^{-r}$. Let $X_j \sim \mathrm{NB}(r_j, p)$, $j = 1, 2, \ldots, m$, be m independent negative binomial random variables. Then, the characteristic function of the sum $S = X_1 + X_2 + \cdots + X_m$ is

$$\begin{aligned}
\phi_S(t) &= \mathbb{E}\left(e^{it(X_1+X_2+\cdots+X_m)}\right) \\
&= \mathbb{E}\left(e^{itX_1} e^{itX_2} \cdots e^{itX_m}\right) \\
&= \mathbb{E}\left(e^{itX_1}\right) \mathbb{E}\left(e^{itX_2}\right) \cdots \mathbb{E}\left(e^{itX_m}\right), \quad (X_j\text{'s are independent}) \\
&= p^{r_1+r_2+\cdots+r_m}(1-(1-p)e^{it})^{-(r_1+r_2+\cdots+r_m)}. \qquad (1)
\end{aligned}$$

The convolution of probability distributions corresponds to the sum of two or more independent random variables. The convolution property of the negative binomial distribution is evident from (1), *i.e.*, $S \sim \mathrm{NB}(r_1 + r_2 + \cdots + r_m, p)$. These results are standard and can be found in any book on basic probability theory.

A probabilistic approach to some identities

We first prove an important identity using the convolution property of the negative binomial distribution.

Math. Mag. **90** (2017) 221–224. doi:10.4169/math.mag.90.3.221.
MSC: Primary 60C05, Secondary 05A19

Theorem 1. *For any positive integers $r_1, r_2, \ldots, r_m$ and a nonnegative integer n, the following identity holds:*

$$\sum_{\Lambda_m^n} \binom{k_1+r_1-1}{k_1}\binom{k_2+r_2-1}{k_2}\cdots\binom{k_m+r_m-1}{k_m} = \binom{n+r-1}{n}, \tag{2}$$

where $r = r_1 + r_2 + \cdots + r_m$, and the sum is taken over the set Λ_m^n consisting of all solutions in nonnegative integers k_j's of $k_1 + k_2 + \cdots + k_m = n$.

Proof. Let $S = X_1 + X_2 + \cdots + X_m$ be the sum of independent negative binomial random variables, *i.e.*, $X_j \sim \text{NB}(r_j, p)$, $j = 1, 2, \ldots, m$. Then,

$$\begin{aligned}
\mathbb{P}\{S = n\} &= \sum_{\Lambda_m^n} \mathbb{P}\{X_1 = k_1, X_2 = k_2, \ldots, X_m = k_m\} \\
&= \sum_{\Lambda_m^n} \mathbb{P}\{X_1 = k_1\}\mathbb{P}\{X_2 = k_2\}\cdots\mathbb{P}\{X_m = k_m\} \qquad (3) \\
&= p^r(1-p)^n \sum_{\Lambda_m^n} \binom{k_1+r_1-1}{k_1}\binom{k_2+r_2-1}{k_2}\cdots\binom{k_m+r_m-1}{k_m}, \qquad (4)
\end{aligned}$$

where (3) follows because the X_j's are independent. Also, since $S \sim \text{NB}(r, p)$, we have

$$\mathbb{P}\{S = n\} = \binom{n+r-1}{n} p^r(1-p)^n. \tag{5}$$

The proof is complete on equating (4) and (5). ■

If we proceed with independent geometric random variables in the proof of the above theorem, *i.e.*, on substituting $r_1 = r_2 = \cdots = r_m = 1$ in (2), the following result is obtained as a special case.

Corollary. For any nonnegative integer n, the total number of nonnegative integer solutions of $k_1 + k_2 + \cdots + k_m = n$ is $\binom{n+m-1}{n}$.

In recent decades, many authors provided probabilistic proofs of certain analytical results such as the fundamental theorem of algebra, Weierstrass approximation theorem, Stirling's formula, etc. For probabilistic proofs of some classical theorems, we refer to [**1**, pp. 11-19]. This includes three different probabilistic proofs of the fundamental theorem of algebra, which states that every polynomial equation of degree $n \geq 1$ with complex coefficients have at least one complex root.

Recently, Kataria [**3**] provided a probabilistic proof of the multinomial theorem using the multinomial distribution. Here, we obtain an alternate probabilistic proof by using the convolution property of the Poisson distribution. A random variable Y is said to have the Poisson distribution with parameter $\lambda > 0$, denoted by $Y \sim \text{P}(\lambda)$, if its probability mass function is given by

$$\mathbb{P}\{Y = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

The sum $T = Y_1 + Y_2 + \cdots + Y_m$ of independent Poisson random variables $Y_j \sim \text{P}(\lambda_j)$ is distributed as $T \sim \text{P}(\lambda_1 + \lambda_2 + \cdots + \lambda_m)$. The proof follows similar lines to that of the convolution property of the negative binomial distribution.

Next, we state the multinomial theorem and prove it using the Poisson distribution.

Theorem 2. (Multinomial Theorem) *Let n be any nonnegative integer and λ_1, $\lambda_2, \dots, \lambda_m$ be real numbers. Then*

$$(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n = \sum_{\Lambda_m^n} \frac{n!}{k_1!k_2!\cdots k_m!} \lambda_1^{k_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m}, \tag{6}$$

where 0^0 is interpreted as unity, and the sum is taken over the set Λ_m^n consisting of all solutions in nonnegative integers k_j's of $k_1 + k_2 + \cdots + k_m = n$.

Proof. Let $T = Y_1 + Y_2 + \cdots + Y_m$ be the sum of independent Poisson random variables, *i.e.*, $Y_j \sim \mathrm{P}(\lambda_j)$, where $\lambda_j > 0$ is assumed for all $j = 1, 2, \dots, m$. Then,

$$\begin{aligned}
\mathbb{P}\{T = n\} &= \sum_{\Lambda_m^n} \mathbb{P}\{Y_1 = k_1, Y_2 = k_2, \dots, Y_m = k_m\} \\
&= \sum_{\Lambda_m^n} \mathbb{P}\{Y_1 = k_1\}\, \mathbb{P}\{Y_2 = k_2\} \cdots \mathbb{P}\{Y_m = k_m\} \qquad (7) \\
&= e^{-(\lambda_1+\lambda_2+\cdots+\lambda_m)} \sum_{\Lambda_m^n} \frac{\lambda_1^{k_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m}}{k_1!k_2!\cdots k_m!}, \qquad (8)
\end{aligned}$$

where (7) follows because Y_j's are independent. Also, since $T \sim \mathrm{P}(\lambda_1 + \lambda_2 + \cdots + \lambda_m)$, we have

$$\mathbb{P}\{T = n\} = \frac{e^{-(\lambda_1+\lambda_2+\cdots+\lambda_m)} (\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n}{n!}. \tag{9}$$

On equating (8) and (9), we have for positive λ_j's

$$(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n = \sum_{\Lambda_m^n} \frac{n!}{k_1!k_2!\cdots k_m!} \lambda_1^{k_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m}, \tag{10}$$

and thus, the multinomial theorem is established for positive reals. Now the proof for the general case, *i.e.*, for all real numbers, follows similar lines to that of Kataria [**3**]. By using the distributive property and on term-by-term multiplication in the left-hand side of (6), we have for all real numbers λ_j's

$$(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n = \sum_{\Lambda_m^n} C(n, k_1, k_2, \dots, k_m) \lambda_1^{k_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m}, \tag{11}$$

where $C(n, k_1, k_2, \dots, k_m)$ are positive integers. On subtracting (10) from (11), we obtain

$$\sum_{\Lambda_m^n} \left(C(n, k_1, k_2, \dots, k_m) - \frac{n!}{k_1!k_2!\cdots k_m!} \right) \lambda_1^{k_1} \lambda_2^{k_2} \cdots \lambda_m^{k_m} = 0, \quad \lambda_j > 0,$$

which is a zero polynomial in m variables, and hence

$$C(n, k_1, k_2, \dots, k_m) = \frac{n!}{k_1!k_2!\cdots k_m!}. \tag{12}$$

This completes the proof. ■

Remark. The number of terms in the multinomial expansion of $(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n$ is precisely the number of nonnegative integer solutions of $k_1 + k_2 + \cdots + k_m = n$, which by the corollary is $\binom{n+m-1}{n}$.

Acknowledgments The author wishes to thank the editor and two anonymous referees for providing insightful comments and suggestions. The research of the author was supported by a UGC fellowship no. F.2-2/98(SA-I), Govt. of India.

REFERENCES

1. K. Burdzy, *Brownian Motion and Its Applications to Mathematical Analysis*, Lecture Notes in Mathematics, Springer, Cham, Switzerland, 2014.
2. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, Third ed., John Wiley & Sons, New York–London–Sydney, 1968.
3. K. K. Kataria, A probabilistic proof of the multinomial theorem, *Amer. Math. Monthly* **123** no. 1 (2016) 94–96.

Summary. We give a simple probabilistic proof of an important combinatorial identity. In the process, we show via probabilistic arguments that there are exactly $\binom{n+m-1}{n}$ terms in the multinomial expansion of $(\lambda_1 + \lambda_2 + \cdots + \lambda_m)^n$. Also, an alternate probabilistic proof of the multinomial theorem is obtained using the convolution property of the Poisson distribution.

KULDEEP KUMAR KATARIA (MR Author ID: 1141643) received his B.Sc. Hons. degree in mathematics from St. Stephen's College, Delhi University. Later, he received his M.Sc. degree in mathematics from the Indian Institute of Technology Kanpur. He is currently pursuing his Ph.D in mathematics from the Indian Institute of Technology Bombay. His research interests lie primarily in fractional calculus and its applications to stochastic processes.

Real Mathematics Fact

Snapple includes "Real Facts" on the underside of the caps to its beverages. A number of them have been about mathematics, including the following.

Snapple's "Real Fact" #804 states that, "There are 293 ways to make change for a dollar." Using pennies, nickels, dimes, quarters, half dollars, and dollar coins, the number of ways to make change for a dollar is equal to the coefficient of the x^{100} term in the series expansion of the closed-form generating function

$$C(x) = \frac{1}{(1-x^1)(1-x^5)(1-x^{10})(1-x^{25})(1-x^{50})(1-x^{100})}.$$

This coefficient is indeed 293.

—Submitted by Stanley R. Huddy,
Fairleigh Dickinson University

Is This the Easiest Proof That nth Roots are Always Integers or Irrational?

JEFFREY BERGEN
DePaul University
Chicago, Illinois 60614
jbergen@depaul.edu

David M. Bloom's *A One-Sentence Proof That* $\sqrt{2}$ *Is Irrational* appeared in this MAGAZINE in 1995 [**1**]. It shows that $\sqrt{k}$ is irrational, whenever $\sqrt{k}$ is not an integer.

There are countless proofs that if $a, n \in \mathbb{N}$, then $\sqrt[n]{a}$ is irrational, provided $\sqrt[n]{a}$ is not an integer. Bloom's article led me to wonder what is the easiest possible proof of this familiar statement. Most proofs of this statement involve concepts such as prime numbers, relative primeness, or the fundamental theorem of arithmetic. However, below is an argument which uses only the well ordering principle and the division algorithm. You can decide if it is the easiest possible proof.

Theorem. *If $a, n \in \mathbb{N}$, then $\sqrt[n]{a}$ is either an integer or is irrational.*

Proof. We need to show that whenever $\sqrt[n]{a}$ is rational, it must be an integer. If $\sqrt[n]{a} \in \mathbb{Q}$, we can write $\sqrt[n]{a} = s/t$ where $s, t \in \mathbb{N}$ and t is the smallest positive denominator occurring in any fraction equal to $\sqrt[n]{a}$. We need to show that t divides s.

Since

$$a = \frac{s^n}{t^n} \quad \text{and} \quad at^n = s^n,$$

we see that t divides s^n. Therefore, there exists a smallest positive integer k such that t divides s^k. If suffices to show that $k = 1$.

By way of contradiction, if $k > 1$, let q, r be the quotient and remainder obtained when s^{k-1} is divided by t. Since s^{k-1} is not a multiple of t, it follows that $t > 1$ and $s^{k-1} = q \cdot t + r$, where $0 < r < t$.

Since s^k is a multiple of t, $s(s^{k-1} - q \cdot t)$ is also a multiple of t. This means that there exists $w \in \mathbb{N}$ such that

$$s \cdot r = s(s^{k-1} - q \cdot t) = t \cdot w \quad \text{so that} \quad \frac{s}{t} = \frac{w}{r}.$$

However $r < t$, contradicting the minimality of t. As a result, $k = 1$; hence t divides s and $\sqrt[n]{a} = \frac{s}{t}$ is an integer. ■

REFERENCE

1. D. M. Bloom, A one-sentence proof that $\sqrt{2}$ is irrational, *Math. Mag.* **68** no. 4 (1995) 286.

Summary. Most proofs that nth roots of positive integers are always integers or irrational involve concepts such as prime numbers, relative primeness, or the fundamental theorem of arithmetic. We provide an argument which uses only the well ordering principle and the division algorithm. You can decide if it is the easiest proof possible.

JEFFREY BERGEN (MR Author ID: 191461) received his B.S. degree from Brooklyn College and his M.S. and Ph.D. degrees from the University of Chicago. He is currently a professor at DePaul University. His research is in noncommutative ring theory and he is the author of the book, "A Concrete Approach to Abstract Algebra: From the Integers to the Insolvability of the Quintic."

Math. Mag. **90** (2017) 225. doi:10.4169/math.mag.90.3.225. © Mathematical Association of America
MSC: Primary 11A05

ACROSS

1. Math subject in H.S.
5. They're sometimes on bulletin boards
10. Catch a few Z's
13. Milky-white gem that's October's birthstone
14. Apple's web browser
15. Fleischer, White House press secretary for George W. Bush
16. * National Mathematics Honor Society that promotes MathFest
18. Former country music TV channel
19. Twitter bot released by Microsoft in 2016 for only 16 hours because it famously devolved into posting offensive tweets
20. *Math Horizons* editor who will give the MAA invited address for students
22. Works at Desmos and will give an invited address entitled "Math's Other Half"
25. Journal paper reviewer, for short
26. Sch. where Terence Tao works
27. Took time away from, for instance, the next scheduled talk
29. Meteor that explodes in the atmosphere
31. Part of a play
32. 2π
34. Suffix with expert
35. * Horne from Morehouse College will give an invited address about his role as a math consultant for this recent film
41. ___ and hers
42. Mathematician consultant for "The Man Who Knew Infinity"
43. One equals about 57.3 degs.
45. "A for ___"
48. * Home of MathFest 2017
51. Narrow opening
52. Form of structural equation modeling used to find subtypes of cases in multivariate categorical data: Abbr.
54. If G is a group, H is a subgroup, and $g \in G$, then $\{gh : h \in H\}$ is this
55. * Williams of Harvey Mudd College who will give an invited address about "unveiling mathematical talent"
58. Member of Congress: Abbr.
59. Mimic
60. Computer science term for, essentially, functions within a larger program
65. Ink on your arm, for short
66. 2001 French romantic comedy film with Audrey Tautou in the title role
67. Bronte heroine
68. Dash widths
69. Hoist again, as a sail
70. * McDuff of Barnard College who will give a series of three invited addresses about symplectic geometry

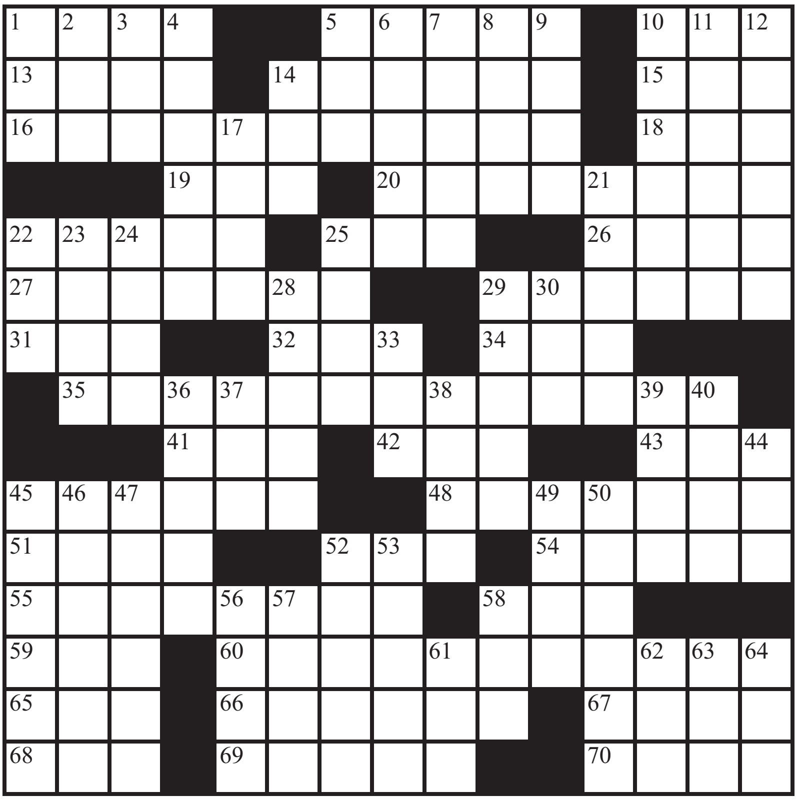
DOWN

1. Logical symbol for a tautology
2. N.Y. engineering sch.
3. Dr. Seuss's Sam ___
4. Posterior muscles
5. Grad students, often
6. Blazing
7. State where 26-Across is: Abbr.
8. Former McDonald's owner Ray, subject of a 2016 Michael Keaton film
9. $\dfrac{e^x - e^{-x}}{2}$
10. Phys., chem., or biol.
11. * Douglas from U. Minnesota who will give an invited address entitled "Computational Math Meets Geometry"
12. Latin term for the leaves of a fern that appear on both sides of a common axis
14. 007, for one
17. Make, an income
21. Swiss mathematician who first solved the Basel problem
22. * Org. behind MathFest
23. Carve in stone
24. Himalayan legend
25. Horse coloring
28. Statistical tool used examine two populations
29. Notation to describe asymptotics, phonetically
30. The Buckeyes, briefly
33. Roswell sighting
36. Traditional garment for Indian men
37. Chairman of the board: Abbr.
38. Ancient civilization that used knots on strings to represent numbers
39. Pitching stats
40. ___Math software
44. It comes before edu
45. ___ sale
46. * Erica from Pomona College who will give an invited address about spatial graph theory
47. French steak cuts, or laces with a square mesh and geometric designs
49. Rapper turned *Law & Order* actor
50. Plagiarized
52. Fibonacci's book: "___ abaci"
53. U.S. women's soccer great Lloyd
56. Ivan the Terrible, e.g.
57. Philosopher David
58. "You will ___ the day!"
61. Part of the Dept. of Health & Human Services
62. * Sch. where Steven Brams works; he'll be giving an invited address about mathematics and democracy
63. Hosp. areas
64. Red, Yellow, or Black ___

Math. Mag. **90** (2017) 226–227. doi:10.4169/math.mag.90.3.226.

MathFest 2017

BRENDAN SULLIVAN
Emmanuel College
Boston, MA
sullivanb@emmanuel.edu



Clues start at left, on page 226. The solution is on page 230.

Extra copies of the puzzle can be found at the MAGAZINE's website, www.maa.org/mathmag/supplements.

Crossword Puzzle Creators

If you are interested in submitting a mathematically themed crossword puzzle for possible inclusion in MATHEMATICS MAGAZINE, please contact the editor at mathmag@maa.org.

The Maclaurin Inequality and Rectangular Boxes

CLAUDI ALSINA
Universitat Politècnica de Catalunya Barcelona, Spain
claudio.alsina@upc.edu

ROGER B. NELSEN
Lewis & Clark College
Portland, OR
nelsen@lclark.edu

In a recent article [**1**], Ben-Ari and Conrad introduced (or reintroduced) readers of this MAGAZINE to Maclaurin's inequality for n nonnegative numbers. A special case is $n = 3$, when the Maclaurin inequality for nonnegative numbers x, y, and z is

$$\frac{x+y+z}{3} \geqslant \sqrt{\frac{xy+xz+yz}{3}} \geqslant \sqrt[3]{xyz} \tag{1}$$

with equality throughout if and only if $x = y = z$. As noted in [**1**], the arithmetic mean–geometric mean inequality for three nonnegative numbers is (1) without the middle term.

Let x, y, and z be the dimensions of a rectangular box (parallelepiped or cuboid). Three quantities of interest associated with the box are its volume $V = xyz$, the total area $F = 2(xy + xz + yz)$ of its six faces, and the total length $E = 4(x + y + z)$ of its 12 edges. Thus, (1) relates these three numbers as

$$\frac{E}{12} \geqslant \sqrt{\frac{F}{6}} \geqslant \sqrt[3]{V} \tag{2}$$

with equality throughout if and only if the box is a cube.

In this note, we present a visual proof of (1) using areas of squares and rectangles and volumes of cubes and boxes after rewriting (1) as

$$(x+y+z)^2 \geqslant 3(xy+xz+yz) \quad \text{and} \quad xy+xz+yz \geqslant 3(xyz)^{2/3}. \tag{3}$$

We assume without loss of generality that $x \geqslant y \geqslant z$ and begin with a simple lemma.

Lemma 1. *If x, y, and z are nonnegative, then $x^2 + y^2 + z^2 \geqslant xy + xz + yz$.*

Proof. See Figure 1, where we compare the areas of three squares to the areas of three rectangles. ∎

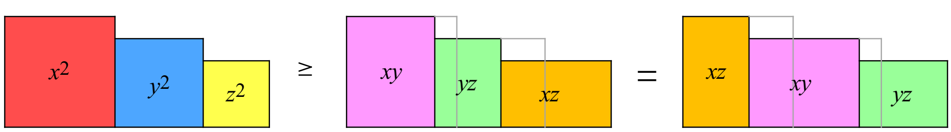


Figure 1

Math. Mag. **90** (2017) 228–230. doi:10.4169/math.mag.90.3.228. © Mathematical Association of America
MSC: Primary 26D07

Theorem 1. *If x, y, and z are nonnegative, then* $(x+y+z)^2 \geqslant 3(xy+xz+yz)$.

Proof. See Figure 2, where we use Lemma 1 to establish the inequality. ■

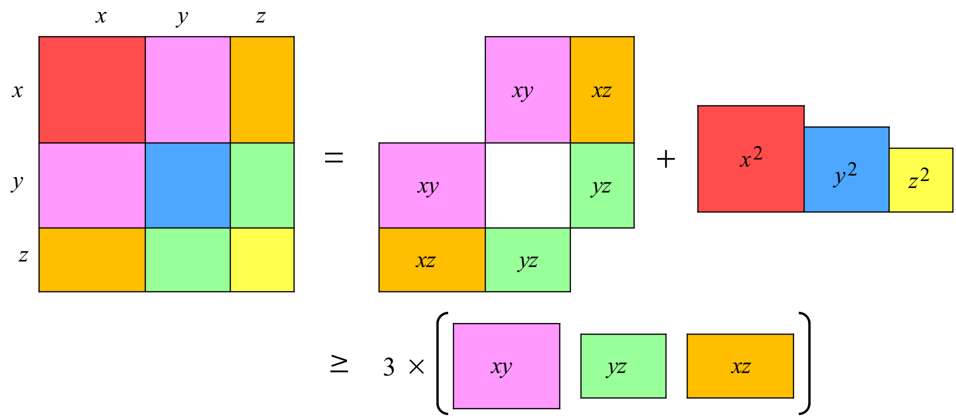


Figure 2

If we let $a^3 = xy$, $b^3 = xz$, and $c^3 = yz$, then $abc = (xyz)^{2/3}$ so that the second inequality in (3) is equivalent to $a^3 + b^3 + c^3 \geqslant 3abc$, which we now prove.

Theorem 2. *If a, b, and c are nonnegative, then* $a^3 + b^3 + c^3 \geqslant 3abc$.

Proof. Let $a \geqslant b \geqslant c$. Comparing volumes of cubes and boxes in Figure 3 (a three-dimensional version of Figure 1), part (a) shows that $a^3 + b^3 + c^3 \geqslant a^2b + b^2c + c^2a$ and part (b) shows that $a^3 + b^3 + c^3 \geqslant a^2c + b^2a + c^2b$. The same two inequalities hold for other orders of a, b, and c.

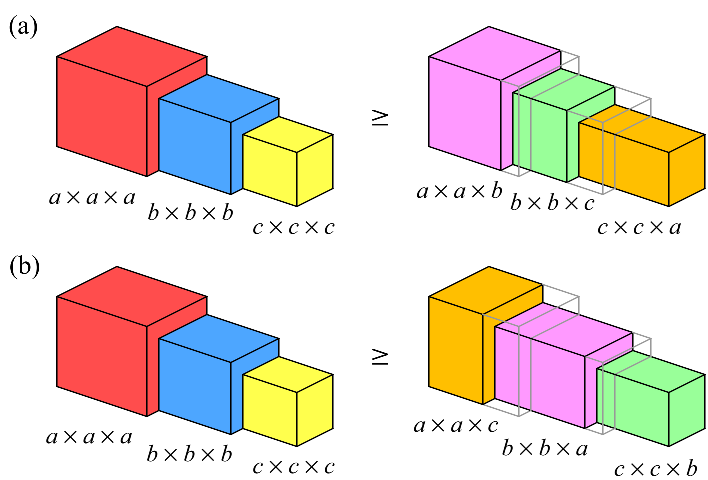


Figure 3

Adding the two inequalities yields

$$2(a^3 + b^3 + c^3) \geqslant a^2b + b^2c + c^2a + a^2c + b^2a + c^2b.$$

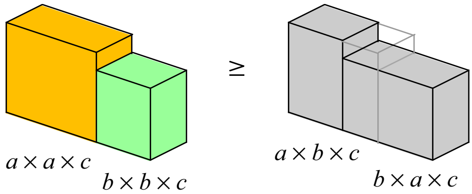


Figure 4

In Figure 4, again comparing volumes, we see that $a^2c + b^2c \geqslant 2abc$, and similar figures show that $a^2b + bc^2 \geqslant 2abc$ and $ab^2 + ac^2 \geqslant 2abc$.

Hence, we have $2(a^3 + b^3 + c^3) \geqslant 6abc$, proving Theorem 2. ■

Lemma 1 and a figure similar to Figure 2 can be used to establish the inequality $3(x^2 + y^2 + z^2) \geqslant (x + y + z)^2$, from which it follows that, if $d = \sqrt{x^2 + y^2 + z^2}$ is the length of the space diagonal of the rectangular box, then $d/\sqrt{3} \geq E/12$.

There are interpretations similar to (2) of the terms of Maclaurin's inequality for $n \geqslant 4$ in terms of n-dimensional volumes of *hyperrectangles* (or *n-orthotopes*) and their lower-dimensional facets. We leave the details to the interested reader.

REFERENCE

1. I. Ben-Ari, K. Conrad, Maclaurin's inequality and a generalized Bernoulli inequality, *Math. Mag.*, **87** (2014), 14–24.

Summary. We interpret the terms of Maclaurin's inequality for three nonnegative numbers as the volume, total face area, and total side length of a rectangular box and provide a visual proof of the inequality.

CLAUDI ALSINA (MR Author ID 25110) is professor of mathematics at Universitat Politècnica de Catalunya.

ROGER B. NELSEN (MR Author ID 237909) is professor emeritus at Lewis & Clark College, where he taught mathematics and statistics for 40 years.

T	R	I	G	■	■	T	A	C	K	S	■	N	A	P
O	P	A	L	■	S	A	F	A	R	I	■	A	R	I
P	I	M	U	E	P	S	I	L	O	N	■	T	N	N
■	■	■	T	A	Y	■	R	I	C	H	E	S	O	N
M	E	Y	E	R	■	R	E	F	■	■	U	C	L	A
A	T	E	I	N	T	O	■	■	B	O	L	I	D	E
A	C	T	■	■	T	A	U	■	I	S	E	■	■	■
■	H	I	D	D	E	N	F	I	G	U	R	E	S	■
■	■	■	H	I	S	■	O	N	O	■	■	R	A	D
E	F	F	O	R	T	■	■	C	H	I	C	A	G	O
S	L	I	T	■	■	L	C	A	■	C	O	S	E	T
T	A	L	I	T	H	I	A	■	R	E	P	■	■	■
A	P	E	■	S	U	B	R	O	U	T	I	N	E	S
T	A	T	■	A	M	E	L	I	E	■	E	Y	R	E
E	N	S	■	R	E	R	I	G	■	■	D	U	S	A

PROBLEMS

Proposals

To be considered for publication, solutions should be received by November 1, 2017.

2021. *Proposed by Mihai Caragiu, Ohio Northern University, Ada, OH.*

For an integer $n > 1$, let $\Pi(n)$ be the greatest prime factor of n. Consider a binary operation '$*$', on the set $\mathcal{P} = \{2, 3, 5, 7, \dots\}$ of all primes, defined by $p * q = \Pi(2p + q)$ for all primes p, q. Find all distinct primes p, q such that $p * q = q * p$.

2022. *Proposed by Mihàly Bencze, Bucharest, Romania.*

Given an irrational number $\beta > 1$, show that there exists a number $\alpha \in (1, 2)$, such that

$$0 < \{\alpha\beta^n\} < \frac{1}{\beta - 1} \quad \text{for all } n \in \mathbb{N},$$

where $\{x\}$ denotes the fractional part of x.

2023. *Proposed by Mircea Merca, Craiova, Romania.*

For every natural number n, let $f(n)$ be the number of representations of n in the form

$$n = a + a + b + c,$$

where a, b, c are distinct positive integers such that $b < c$. Show that there are infinitely many values of n such that $f(n + 1) < f(n)$.

Math. Mag. **90** (2017) 231–238. doi:10.4169/math.mag.90.3.231. © Mathematical Association of America

We invite readers to submit original problems appealing to students and teachers of advanced undergraduate mathematics. Proposals must always be accompanied by a solution and any relevant bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. Submitted problems should not be under consideration for publication elsewhere.

Proposals and solutions should be written in a style appropriate for this MAGAZINE.

Authors of proposals and solutions should send their contributions using the Magazine's submissions system hosted at http://mathematicsmagazine.submittable.com. *More detailed instructions are available there. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* mathmagproblems@maa.org.

2024. *Proposed by George Stoica, Saint John, New Brunswick, Canada.*

A *binary expansion* is an expression of the form

$$0.d_1d_2d_3\cdots d_i\cdots,$$

where each numeral (digit) d_i is either 0 or 1 ($i = 1, 2, 3, \ldots$). Given a real number $\beta > 1$ (called the *base*) the *base-β value* of the binary expansion above is

$$(0.d_1d_2d_3\cdots)_\beta = \sum_{i=1}^{\infty} \frac{d_i}{\beta^i}.$$

(i) If $1 < \beta < 2$, show that infinitely many distinct binary expansions have base-β value equal to 1.

(ii) Find all binary expansions with value 1 when $\beta = (1+\sqrt{5})/2$ (the golden ratio).

2025. *Proposed by Valerian Nita, Sterling Heights, MI.*

Let n be a positive integer and let $x_1, x_2, \ldots, x_n$ and $a_1, a_2, \ldots, a_n$ be real numbers such that $\sum_{k=1}^{n} x_k = 0$ and $0 < a_1 < a_2 < \cdots < a_n$. Define $s_1, s_2, \ldots, s_n$ by $s_k = \sum_{j=1}^{k} a_j x_j$ for $k = 1, 2, 3, \ldots, n$. If there is at least one nonzero number among $x_1, x_2, \ldots, x_n$, prove that there is at least one positive and at least one negative number among $s_1, s_2, \ldots, s_n$.

Quickies

1071. *Proposed by Leonard Giugiuc (Romania) and Kadir Altintaş (Turkey).*

Let G be the centroid of triangle $\triangle ABC$, D be the midpoint of $\overline{BC}$, and E the midpoint of $\overline{AC}$. Prove that the quadrilateral $DCEG$ is cyclic if and only if $AC^2 + BC^2 = 2AB^2$.

1072. *Proposed by John Zacharias, Arlington, VA.*

For every natural number n, prove the trigonometric identity

$$\sum_{k=1}^{n} \frac{\sin\theta}{\cos\left(\frac{2\pi k}{n}\right) - \cos\theta} = n\cot\frac{n\theta}{2}.$$

(When the expression on either side is undefined, the identity states that so is the one on the other side.)

Solutions

A limit of sums of reciprocals of quartics **April 2016**

1991. *Proposed by George Stoica, Saint John, New Brunswick, Canada.*

Evaluate

$$\lim_{n\to\infty} \sum_{k=1}^{n-1} \frac{n^2}{k^2(n-k)^2}.$$

Solution by Eugene A. Herman, Grinnell College, Grinnell, IA.
We have the partial-fraction decomposition

$$\frac{n^2}{k^2(n-k)^2} = \frac{2}{nk} + \frac{1}{k^2} + \frac{2}{n(n-k)} + \frac{1}{(n-k)^2}.$$

Hence,

$$\sum_{k=1}^{n-1} \frac{n^2}{k^2(n-k)^2} = \frac{2}{n}\sum_{k=1}^{n-1}\left(\frac{1}{k} + \frac{1}{n-k}\right) + \sum_{k=1}^{n-1}\left(\frac{1}{k^2} + \frac{1}{(n-k)^2}\right)$$
$$= \frac{4}{n}H_{n-1} + 2\sum_{k=1}^{n-1}\frac{1}{k^2},$$

where $H_n = \sum_{k=1}^{n} k^{-1}$ is the nth harmonic number. Since $H_n = \log n + \gamma + O(1/n)$ as $n \to \infty$ and $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$,

$$\lim_{n\to\infty} \sum_{k=1}^{n-1} \frac{n^2}{k^2(n-k)^2} = 2\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{3}.$$

Also solved by Adnan Ali (India), Michel Bataille (France), Elton Bojaxhiu (Albania) & Enkel Hysnelaj (Australia), Brian Bradie, Robert Calcaterra, Robin Chapman (UK), Hongwei Chen, Richard Daquila, Dionne Bailey & Elsie Campbell & Charles Diminnie, Natacha Fontes-Merz, Michael Goldenberg & Mark Kaplan, Jeffrey M. Groah, GWstat Problem Solving Group, Tom Jager, Koopa Tak-Lun Koo (Hong Kong), Kee-Wai Lau (China), Clarence W. Lienhard, James Magliano, John Mahony (New Zealand), Matthew McMullen, Jerry Minkus, Missouri State University Problem Solving Group, Rituraj Nandan, Roger Peterson, Rob Pratt, Henry Ricardo, Jim Vandergriff, Michael Vowe (Switzerland), Edward T. White, John Zacharias, and the proposer. There was 1 incomplete or incorrect solution.

Functions with infinite limit at all rationals **April 2016**

1992. *Proposed by Oniciuc Gheorghe, Botosani, Romania.*

(a) Show that there is no function $f : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to a} |f(x)| = \infty$ for every $a \in \mathbb{Q}$.

(b) Show that there exists a function $f : \mathbb{R} \to [-\infty, +\infty]$ such that f is finite almost everywhere and $\lim_{x\to a} f(x) = +\infty$ for every $a \in \mathbb{Q}$.

Solution by Robin Chapman, Department of Mathematics, University of Exeter, UK.

(a) Suppose $f : \mathbb{R} \to \mathbb{R}$ has the property that $\lim_{x\to a} |f(x)| = \infty$ for each $a \in \mathbb{Q}$. For each positive integer n, define

$$A_n = \{x \in \mathbb{R} : |f(x)| > n\},$$

and let U_n be the interior of A_n (i.e., the union of all open intervals included in A_n). Since f takes only finite values, we have $\bigcap_{n=1}^{\infty} A_n = \emptyset$, so $\bigcap_{n=1}^{\infty} U_n = \emptyset$ *a fortiori*.

We claim that each U_n is dense in $\mathbb{R}$. To prove this, it suffices to show that $U_n \cap I \neq \emptyset$ for any nonempty open interval $I = (r, s)$. Any such I certainly contains some rational number a. Since $\lim_{x\to a} |f(x)| = \infty$, there exists $\varepsilon > 0$ such that $|f(x)| > n$ when $0 < |x - a| < \varepsilon$; in particular, $J = (a, a + \varepsilon) \subseteq A_n$, hence $J \subseteq U_n$. We have $I \cap J = (a, t)$ with $t = \min\{s, a + \varepsilon\} > a$, so $I \cap J \neq \emptyset$, and $I \cap U_n \neq \emptyset$ *a fortiori*. We conclude that U_n is dense in $\mathbb{R}$.

Since $\mathbb{R}$ is a complete metric space, the Baire category theorem implies that the intersection of any countable sequence of dense open subsets of $\mathbb{R}$ is dense in $\mathbb{R}$, hence nonempty. Therefore, $\bigcap_{n=1}^{\infty} U_n \neq \emptyset$, a contradiction. A function with the properties stated in (a) cannot exist.

(b) Let $(b_n)_{n=1}^{\infty}$ be a sequence of rational numbers taking every rational value infinitely many times. (Such a sequence exists since $\mathbb{Q}$ is countable.) Let $I_n = (b_n - 2^{-(n+1)}, b_n + 2^{-(n+1)})$ be the open interval of length 2^{-n} centered at b_n. Let g_n denote the characteristic function of I_n. For $x \in \mathbb{R}$, define

$$f(x) = \sum_{n=1}^{\infty} g_n(x) \in [0, \infty],$$

thus obtaining a well-defined function $f : \mathbb{R} \to [0, \infty]$. We show that f satisfies the properties stated in part (b) of the problem.

For each $n \geq 1$,

$$\int_{-\infty}^{\infty} g_n(x)\, dx = 2^{-n}.$$

By the monotone convergence theorem, f is a Lebesgue-integrable function with

$$\int_{-\infty}^{\infty} f(x)\, dx = \sum_{n=1}^{\infty} 2^{-n} = 1.$$

Since f is nonnegative with finite Lebesgue integral, it follows that set of x such that $f(x) = \infty$ has Lebesgue measure zero, so f is finite almost everywhere.

Let $a \in \mathbb{Q}$ and let N be a positive integer. Let b_m be the Nth occurrence of a in the sequence (b_n). Then a is the midpoint of N intervals J_k with $k \leq m$, and $J_k \supseteq J_m$ for each of these k. Therefore, $f(x) \geq N$ for $x \in J_m$. It follows that $f(x) \to \infty$ as $x \to a$. This completes the solution.

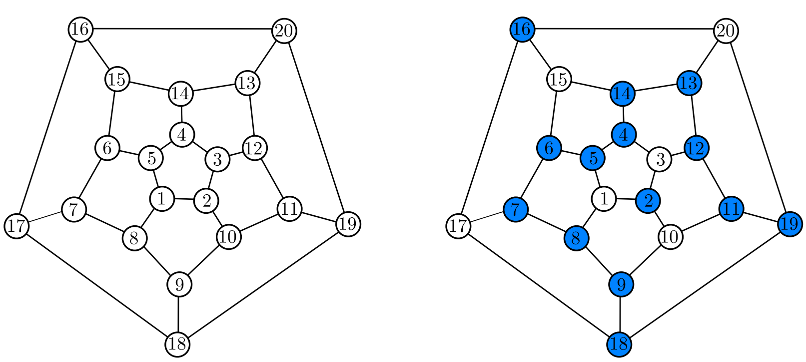
Also solved by Robert Calcaterra, Richard Daquila, Tom Jager, Northwestern University Math Problem Solving Group, José M. Pacheco & Ángel Plaza (Spain), Stephen Scheinberg, Lawrence R. Weill and the proposer.

Coloring the faces of an icosahedron **April 2016**

1993. *Proposed by Kimberly D. Apple, Columbus State University, GA.*

Each face of an icosahedron is colored blue or white in such a way that any blue face is adjacent to no more than two other blue faces. What is the maximum number of blue faces? (Two faces are considered adjacent if they share an edge.)

Solution by Robert Calcaterra, University of Wisconsin-Platteville, Platteville, WI.



The vertices of the graph G above on the left correspond to a possible numbering of the faces of an icosahedron. The edges of G connect adjacent faces of the icosahedron. We will say that a vertex coloring of G using blue and white is *admissible* if any blue vertex is adjacent to no more than two other blue vertices. Admissible colorings of G obviously correspond to colorings of the faces of an icosahedron as in the statement of the problem. Coloring vertices 1, 3, 10, 15, 17, and 20 white and the remaining vertices blue we obtain an admissible coloring (right figure above). Therefore, the maximum number of blue vertices in an admissible coloring of G is at least fourteen.

We show that a coloring with at least fifteen blue vertices is not admissible. For $n = 1, 2, \ldots, 20$, let $Y_n = \{n, a, b, c\}$, where a, b, c are the numbers of the three distinct vertices adjacent to vertex n. (For example, $Y_5 = \{5, 1, 4, 6\}$ and $Y_{17} = \{17, 7, 16, 18\}$.) The sets $Y_1, Y_2, \ldots, Y_{20}$ are obviously all distinct. Assume an admissible coloring C with exactly fifteen blue and five white vertices exists. Since C is admissible, each of the sets Y_n has at least one member colored white. Since each element of $\{1, 2, \ldots, 20\}$ belongs to exactly four of the sets Y_n, it follows that each set Y_n must have exactly one member colored white. By symmetry, we may assume that vertex 1 is colored white without loss of generality. Since 1 belongs to Y_1, Y_2, Y_5, and Y_8, vertices 2 through 10 must all be colored blue. Since three of the members of each of the sets Y_9 and Y_{10} are colored blue, it follows that their fourth members 11 and 18 must be colored white. However, 11 and 18 are both members of Y_{19}, contradicting the assumed admissibility of C. We conclude that an admissible coloring with fifteen blue vertices does not exist. Since any admissible coloring remains admissible if an arbitrary blue vertex is re-colored white, no admissible coloring has more than fifteen blue vertices either. Therefore, the maximum number of blue vertices in an admissible coloring of G (and thus of the faces of the icosahedron in accordance with the conditions of the problem) is fourteen.

Also solved by Michel Bataille (France), Robin Chapman (UK), Joseph DiMuro, Tommy Goebeler & Elias Lindgren & Joshua Wang, Rob Pratt, San Francisco University High School Problem Solving Group, Stan Wagon, and the proposer. There was 1 incomplete or incorrect solution.

A golden random walk in the integers **April 2016**

1994. *Proposed by Donald E. Knuth, Computer Science Department, Stanford University, CA.*

Let $X_0 = 0$, and suppose X_{n+1} is equally likely to be either $X_n + 1$ or $X_n - 2$. What is the probability, p_m, that $X_n \le m$ for all $n \ge 0$?

Solution by Rob Pratt, Washington, DC.
We make the explicit assumption that the increments $\Delta_n = X_{n+1} - X_n$ are independent, and show that

$$p_m = 1 - \left(\frac{\sqrt{5}-1}{2}\right)^{m+1} \quad \text{for } m \ge 0.$$

(Since $X_0 = 0$, we have $p_m = 0$ if $m < 0$.)

Let q_m denote the probability that $X_n \ge m$ for some $n \ge 0$, so that $q_0 = 1$ and $p_m = 1 - q_{m+1}$. Considering the disjoint cases $X_1 = 1$, $X_1 = -2$ separately and conditioning on the value of X_1, we have

$$q_m = \frac{1}{2}q_{m-1} + \frac{1}{2}q_{m+2} \quad \text{for } m \ge 1, \tag{1}$$

since the sequence $\{Y_0, Y_1, \dots\}$ with $Y_n = X_{n+1} - X_1$ has the same distribution as $\{X_0, X_1, \dots\}$, by the independence assumption ($Y_n \geq m - 1$ in the first case, $Y_n \geq m + 2$ in the second, for some $n \geq 0$).

For $m, n \geq 1$, the event $X_n \geq m$ is obviously a subevent of $X_n \geq 1$. If $X_n \geq 1$ for some n, choose n_0 minimal such that $X_{n_0} \geq 1$. Clearly we must have $X_{n_0} - X_{n_0-1} = 1$, so $X_{n_0-1} = 0$ and $X_{n_0} = 1$. Given $m \geq 1$ and the values $X_0 = 0, X_1, X_2, \dots, X_{n_0} = 1$ (with $X_n \leq 0$ for $0 \leq n < n_0$), the conditional probability that $X_n \geq m$ for some n is exactly the probability that $Y_n \geq m - 1$ for some n, where $Y_n = X_{n+n_0} - 1$ for $n \geq 0$. By the independence assumption, the latter probability is equal to q_{m-1}. It follows that $q_m = q_1 q_{m-1}$ for all $m \geq 1$ and, since $q_0 = 1$, induction gives

$$q_m = q_1^m \quad \text{for } m \geq 0. \tag{2}$$

Substituting (2) into (1), we have $q_1^m = (q_1^{m-1} + q_1^{m+2})/2$ for $m \geq 1$; equivalently, $q_1^3 - 2q_1 + 1 = 0$, whose nonnegative roots are 1 and $\phi^{-1} = (\sqrt{5} - 1)/2$. The strong law of large numbers applied to the sequence $\{\Delta_n = X_{n+1} - X_n\}$ implies that, almost surely, $X_n/n \to (1-2)/2 = -1/2$, and hence $X_n \to -\infty$. Thus, the sequence $\{X_0, X_1, \dots\}$ is bounded above almost surely, hence $q_m \to 0$ as $m \to \infty$, so $q_1 < 1$ from (2). We must have $q_1 = \phi^{-1}$, so $p_m = 1 - q_{m+1} = 1 - \phi^{-(m+1)}$ for $m \geq 0$.

Also solved by Robert A. Agnew, Michael Andreoli, Robert Calcaterra, Robin Chapman (UK), Tom Jager, and the proposer. There were 5 incomplete or incorrect solutions.

The limit of powers of a stochastic matrix — April 2016

1995. *Proposed by Michel Bataille, Rouen, France.*

Let $a_1, a_2, \dots, a_n$ be distinct positive real numbers ($n > 2$). For $j = 1, 2, \dots, n$, let $s_j = \sum_{i \neq j} a_i$. Consider the $n \times n$ matrix $M = (m_{i,j})$ defined by $m_{i,i} = 0$ and $m_{i,j} = a_i/s_j$ for $1 \leq i \neq j \leq n$. Show that there exist column vectors X and Y such that

$$X^T Y = 1 \quad \text{and} \quad Y X^T = \lim_{k \to \infty} M^k.$$

Solution by Nicholas C. Singer, Annandale, VA.

(The hypothesis that $a_1, a_2, \dots, a_n$ are distinct is superfluous, as the solution shows.)

The vector $X = (1, 1, \dots, 1)^T$ is a left eigenvector of M with eigenvalue 1, since

$$(X^T M)_i = \sum_{k=1}^n X_k m_{k,i} = \sum_{k=1}^n m_{k,i} = \sum_{k \neq i} \frac{a_k}{s_i} = \frac{s_i}{s_i} = 1 = (X^T)_i.$$

It follows that M is a stochastic matrix since it has nonnegative entries. Moreover, M is primitive, hence irreducible, since the entries of M^2,

$$(M^2)_{i,j} = \sum_{k=1}^n m_{i,k} m_{k,j} = \sum_{k \neq i, k \neq j} \frac{a_i a_k}{s_k s_j},$$

are all positive under the hypothesis $n > 2$. Let $A = \sum_{k=1}^n a_k s_k$ and let Y be the column vector with entries $Y_i = a_i s_i / A$. We have

$$(MY)_i = \sum_{k=1}^n m_{i,k} Y_k = \sum_{k \neq i} \frac{a_i}{s_k} \frac{a_k s_k}{A} = \frac{a_i}{A} \sum_{k \neq i} a_k = \frac{a_i s_i}{A} = Y_i, \quad \text{and}$$

$$X^T Y = \sum_{i=1}^n 1 \cdot \frac{a_i s_i}{A} = \frac{A}{A} = 1.$$

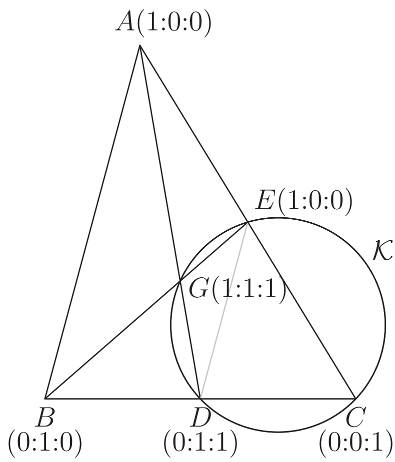
Thus, M has Y as a right eigenvector with eigenvalue 1. By the Perron–Frobenius theorem for primitive stochastic matrices, the eigenvalue 1 for M is simple, and all other eigenvalues have magnitude < 1. The unique matrix of rank 1 whose only nonzero eigenvalue is 1 with corresponding left and right eigenvectors X, Y is the matrix YX^T, so we have $\lim_{k\to\infty} M^k = YX^T$.

Also solved by Mark Kaplan, Jeffrey Stuart, Robert Calcaterra, Eugene Herman, and the proposer. There were 3 incomplete or incorrect solutions.

Answers

Solutions to the Quickies from page 232.

A1071. In the (homogeneous) barycentric coordinate system based on $A(1{:}0{:}0)$, $B(0{:}1{:}0)$ and $C(0{:}0{:}1)$, the remaining points are $D(0{:}1{:}1)$, $E(1{:}0{:}1)$ and $G(1{:}1{:}1)$ as in the figure.



Letting $a = BC$, $b = AC$, $c = AB$, it is well known that the equation of a circle $\mathcal{K}$ in barycentric coordinates $(x{:}y{:}z)$ takes the form

$$a^2yz + b^2xz + c^2xy = (x\mathrm{p}_{\mathcal{K}}(A) + y\mathrm{p}_{\mathcal{K}}(B) + z\mathrm{p}_{\mathcal{K}}(C))(x + y + z),$$

where $\mathrm{p}_{\mathcal{K}}(S)$ denotes the power of an arbitrary point S with respect to $\mathcal{K}$. Specializing to the case when $\mathcal{K}$ is the circumcircle of triangle $\triangle CDE$ we have $p_{\mathcal{K}}(A) = AC \cdot AE = b^2/2$, $p_{\mathcal{K}}(B) = BC \cdot BD = a^2/2$, and $p_{\mathcal{K}}(C) = 0$ since C lies on $\mathcal{K}$. Thus, $\mathcal{K}$ has equation

$$a^2yz + b^2xz + c^2xy = \frac{1}{2}(b^2x + a^2y)(x + y + z).$$

The points C, D, E, G are concyclic precisely when $G(1{:}1{:}1)$ satisfies the equation above, which happens just when $a^2 + b^2 + c^2 = 3(a^2 + b^2)/2$; equivalently, when $a^2 + b^2 = 2c^2$.

A1072. For $k = 1, 2, \ldots, n$, let $\zeta_k = \exp(2\pi ik/n)$ and $c_k = \cos(2\pi k/n)$. The roots of the polynomial $z^n - 1$ are all simple, lying at $z = \zeta_1, \zeta_2, \ldots, \zeta_n$. Since $1/(z^n - 1)$ has residue ζ_k/n at $z = \zeta_k$, we have a partial-fraction decomposition

$$\frac{1}{z^n-1}=\frac{1}{n}\sum_{k=1}^{n}\frac{\zeta_k}{z-\zeta_k}.$$

Let $z=\exp(i\theta)$. For $k=1,2,\ldots,n$, we have

$$\frac{\sin\theta}{\cos(2\pi k/n)-\cos\theta}=\frac{\frac{z-z^{-1}}{2i}}{c_k-\frac{z+z^{-1}}{2}}=i\cdot\frac{z^2-1}{z^2-2c_kz+1}=i\left(1+\frac{\zeta_k}{z-\zeta_k}+\frac{\overline{\zeta}_k}{z-\overline{\zeta}_k}\right).$$

Since $\overline{\zeta}_n=1=\zeta_n$ and $\overline{\zeta}_k=\zeta_{n-k}$ for $1\le k<n$,

$$\sum_{k=1}^{n}\frac{\sin\theta}{\cos(2\pi k/n)-\cos\theta}=ni+2i\sum_{k=1}^{n}\frac{\zeta_k}{z-\zeta_k}=ni\left(1+\frac{2}{z^n-1}\right)$$

$$=ni\cdot\frac{z^{n/2}+z^{-n/2}}{z^{n/2}-z^{-n/2}}=n\cot\frac{n\theta}{2}.$$

Either side of the identity is undefined precisely when $\theta=2\pi k/n$ for an arbitrary integer k.

PINEMI PUZZLE

	5		5			7			
8		9		9	8		9		5
		10			7	6		8	
	10		9			7	8		
3							8		8
6	9			10		10		7	
		10	11		9			8	5
6	10			10			8		
	8		9			10		10	
4		6			6		8		4

How to play. Place one jamb (|), two jambs (||), or three jambs (|||) in each empty cell. The numbers indicate how many jambs there are in the surrounding cells—including diagonally adjacent cells. Each row and each column has 10 jambs. Note that no jambs can be placed in any cell that contains a number.

The solution is on page 181.

—contributed by Lai Van Duc Thinh
Vietnam; fibona2cis@gmail.com

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles, books, and other materials are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Shetterly, Margot Lee, *Hidden Figures: The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race*, William Morrow, 2016; 368 pp, $27.99, $15.99 (P). ISBN 978-0-06236359-6. Excerpted as: The woman the Mercury astronauts couldn't do without, *Nautilus* No. 17 (Nov/Dec 2016) 22–39; http://nautil.us/issue/43/heroes/the-woman-the-mercury-astronauts-couldnt-do-without.

Hidden Figures. DVD, 20th Century Fox, 2017; 127 min, $29.98.

Bryant, Jeffrey, Paco Jain, and Michael Trott, Hidden Figures: Modern approaches to orbit and reentry calculation, http://blog.wolfram.com/2017/02/24/hidden-figures-modern-approaches-to-orbit-and-reentry-calculations/.

By now, you likely have heard of—and perhaps seen—the film "Hidden Figures," particularly since it was nominated as Best Picture for the Academy Awards. Both the film and the book that it was based on relate in both serious and entertaining fashion the struggles and successes of the heretofore unrecognized "colored women computers": They did orbital calculations in the 1950s for the predecessor of NASA, before computers replaced them. (Mathematicians will be amused by the calls for "computers" who know analytic geometry.) The culminating event of the film is the launch and retrieval of John Glenn on the first U.S. human orbital flight. Bryant et al. reproduce calculations by "computer" Katherine Johnson, protagonist of the book and the film, who worked on the Glenn flight, which the authors confirm with computations using tools from the Wolfram language.

Silva, Jorge Nuno (ed.), *Proceedings of the Recreational Mathematics Colloquium IV: Gathering for Gardner Europe*, Lisbon, 2016; iii + 177 pp, $24.21(P) from Amazon.

In odd-numbered years since 2009, there has been a Recreational Mathematics Colloquium in Portugal, complementing the U.S. Gatherings4Gardner in even-numbered years. This full-color volume of proceedings of the "2015 Gathering for Gardner Europe" includes articles on symmetries in Portuguese pavements, negative prices(!) in U.K. supermarkets, mathematics activities for kindergarten, wallpaper patterns for mathematics teachers, a history of knight's tours of the chessboard, and a biography of the author of *Flatland*. The 2017 Gathering has already taken place; the next one is at the University of Coimbra, Portugal, January 26–29, 2019.

Najmabadi, Shannon, Meet the math professor who's fighting gerrymandering with geometry, *Chronicle of Higher Education* (10 March 2017) http://www.chronicle.com/article/Meet-the-Math-Professor/239260.

Moon Duchin (Tufts Univ.) realized that metric geometry can be applied to gerrymandering (manipulating voter district shapes to favor one party). Going beyond theory, she has created a five-day summer program to train mathematicians as expert witnesses in court cases over the drawing of voter district maps (https://sites.tufts.edu/gerrymandr/). "What courts have been looking for is one definition of compactness that they can understand, that we can compute, and that they can use . . . . I think we can make a contribution to the debate."

Math. Mag. **90** (2017) 239–240. doi:10.4169/math.mag.90.3.239. © Mathematical Association of America

Albert, Jim, *Teaching Statistics Using Baseball*, 2nd ed., MAA, 2017; xi + 243 pp, \$55(P), \$41.25 MAA member. ISBN 978-1-93951-216-1.

This new edition updates Albert's 2003 book to use data from current teams and players, together with the addition of new kinds of data about pitches (pitchFX) and runners (Statcast), as well as new metrics and analytics. Statistical topics include the usual descriptive statistics techniques (numeric and graphical), standardization, scatterplots and least-squares lines, regression to the mean, probability, a taste of hypothesis testing, streakiness, and modeling baseball as a Markov chain. The author teaches introductory statistics from this book to an audience "taking the class to satisfy their mathematics requirement." I am conflicted about this aim since, in my experience, introductory statistics is in part a *client discipline* of mathematics, much as introductory physics is; an introductory course in statistics uses first-year algebra but introduces no mathematics that is new to the students. In this book, though, Markov chains and matrices come into play.

Griggs, Terry S., Combinatorics of the sonnet, *Journal of Humanistic Mathematics* 6 (2) (July 2016) 38–46. DOI: 10.5642/jhummath.201602.05. Available at http://scholarship.claremont.edu/jhum/vol6/iss2/5.

A sonnet is a poem in English of 14 lines with one of several specified rhyming schemes. The 14 lines are divided into the first eight (the *octave*, divided into two *quatrains* of four lines each) and the last six (the *sestet*). The author counts the number of possible rhyming schemes, relates them to one another via graphs, and along the way introduces *Tutte's (3,8) cage*. Shakespeare's sonnets all have the same rhyming scheme, and it is not clear that sonnets have been written in all 225 possible basic rhyme forms (author Griggs counts schemes only from sonnets of poet John Clare). Were I more clever, I should have written this review in sonnet form!

Holden, Joshua, *The Mathematics of Secrets: Cryptography from Caesar Ciphers to Digital Encryption*, Princeton University Press, 2017; xiv + 373 pp, \$29.95. ISBN 978-0-691-14175-6.

Bauer, Craig, *Unsolved! The History and Mystery of the World's Greatest Ciphers from Ancient Egypt to Online Secret Societies*, Princeton University Press, 2017; xi + 600 pp, \$35. ISBN 978-0-691-16767-1.

Author Holden describes traditional elements of ciphers systems and breaking them. Modular arithmetic and some probability are needed; permutations, elliptic curves, and lattices appear, too. Included is a "secret history of public-key cryptography" (pp. 235–240) with more details than I had known. Bauer's book contains little mathematics. It begins with almost 100 pp about what is known as the Voynich manuscript, a 15th-century cipher work still unbroken today. Other chapters treat more-ancient ciphers, one by Edward Elgar, some by serial killers, and many other challenges. The only dissatisfaction is that despite tantalizing clues, all the ciphers remain unsolved.

Lane, Matthew, *Power-Up: Unlocking the Hidden Mathematics in Video Games*, Princeton University Press, 2017; xii + 272 pp, \$29.95. ISBN 978-0-691-16151-8.

I remember a time long ago when companies desperately seeking computer programmers were willing to hire people with little or no mathematical background; such a time seems to have come again, as I observe eager hiring of current computer science majors weak in minimal mathematics (calculus, discrete structures) and fearful of, or uninterested in, learning more. Perhaps this book will excite them and pique their curiosity, particularly the large proportion of them seeking to write video games. From the correct physics of Mario jumping out of a plane, through the chances of repeat questions in Super Nintendo's *Family Feud*, to the investigation of Bayesian regret in a multicandidate generalization of Nintendo's *Everybody Votes*, and a model for the highest possible score in *Donkey Kong*, there are hints and examples of lots of mathematics here: the birthday problem, the coupon collector problem, Arrow's impossibility theorem, normal curve probabilities, logistic curves, Apollonian circles, arclength, P vs. NP, and even the Hardy–Ramanujan approximation to the growth rate of the partition function.

MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St., NW
Washington, DC 20036

CONTENTS

ARTICLES

PROBLEMS AND SOLUTIONS

REVIEWS