Vol. 86, No. 2, April 2013

MAA

MATHEMATICS MAGAZINE

A Deranged Sock Monkey (p. 109)

- Deranged gloves vs. deranged socks
- From doodles to diagrams to knots
- Using statistics in experimental physics
- Playing bocce with probabilistic linear algebra

An Official Publication of the **Mathematical Association of America**

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.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 83, at pages 73–74, and is available at the *Magazine*'s website www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Please submit new manuscripts by email directly to the editor at mathmag@maa.org. A brief message containing contact information and with an attached PDF file is preferred. Word-processor and DVI files can also be considered. Alternatively, manuscripts may be mailed to Mathematics Magazine, 132 Bodine Rd., Berwyn, PA 19312-1027. If possible, please include an email address for further correspondence.

Cover image by Annie Stromquist

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Lancaster, PA, bimonthly except July/August.
The annual subscription price for *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to

MAA Advertising
1529 Eighteenth St. NW
Washington DC 20036

Phone: (877) 622-2373
E-mail: tmarmor@maa.org

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

Change of address, missing issue inquiries, and other subscription correspondence:

MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

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, N.W., Washington, D.C. 20036-1385.

Printed in the United States of America

Vol. 86, No. 2, April 2013

MAA

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

What is special about the doodles on the facing page? For one thing, they are the starting point for an excellent adventure in graph theory and knot theory, guided by master tour guides Colin Adams, Nöel MacNaughton, and Charmaine Sia. Their article is based largely on two of the authors' undergraduate research projects.

Sally Cockburn and Joshua Lesperance tell us about the "deranged sock problem." Two socks are given to each of n people. The socks form n pairs, each with a distinctive color, but they are handed out without regard to the colors. What is the probability that at least one person receives a matching pair? I posed a version of this problem to students in a counting class. They rose to the challenge and ran amazingly far with it. But these authors have taken it much further, and my next class will benefit from their methods.

The next article shows statistics in the service of experimental physics—in examples involving spinning disks, gamma-ray counters, and the cosmic background radiation. The examples are interesting in themselves, but the five authors admit to an ulterior motive: They hope that these examples, and other examples from physics, will be used in statistics classes, enriching the teaching of both physics and statistics.

Kent Morrison puts probability in the service of a lunchtime bocce game. But he doesn't stop there. Suppose that you are trying to construct a finite probability distribution. You have built a system of linear equations, and you want your vector of probabilities to be a solution of the linear system. Will your plan succeed? Maybe—but only if the system has a solution with all positive entries. For a random linear system, what is the probability this will occur? The answer is found by the same methods that applied to the bocce game.

A function can be defined on $[0, 1]$ and differentiable only at the irrationals. You may know an example. But did you know that a function can be continuous on $[0, 1]$ and differentiable only at the rationals? Mark Lynch gives us the definitive example in the Notes Section.

Also in the Notes, Sam Northshield has a fresh approach to Marsden's theorem—the one about the critical points of a cubic polynomial being the foci of a certain ellipse. Remember what you read about this theorem, because we have other articles about it in the pipeline.

Finally, Sunil Chebolu and Michael Mayers answer one more question: What is special about the divisors of 12?

Walter Stromquist, Editor

ARTICLES

From Doodles to Diagrams to Knots

COLIN ADAMS
Williams College
Williamstown, MA 01267
Colin.C.Adams@williams.edu

NÖEL MacNAUGHTON
New York, NY
noel.macnaughton@bankofamerica.com

CHARMAINE SIA
Harvard University
Cambridge, MA 02138
sia@math.harvard.edu

You are sitting in a colloquium entitled "Bifurcational Operators in Monoidally Eviscerated Cotomology." About halfway into it, you become hopelessly lost, so you begin to doodle. Without lifting your pen from the page, you draw a curve that intersects itself and closes up, as in FIGURE 1. You immediately notice a peculiarity of the drawing. All of the complementary regions defined by the curve, those faces that would result if the curve were removed from the plane, including the outside one, have either two or five sides. "Cool," you think. "I wonder if I can draw another picture for which that is true."

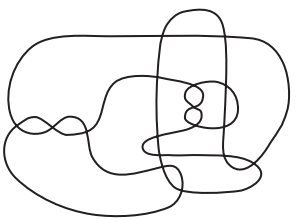


Figure 1 A doodle with only bigons and pentagons as the complementary regions

So you continue to doodle. After quite a bit of trial and error, you produce the curve in FIGURE 2.

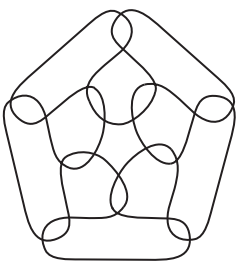


Figure 2 Another doodle with only bigons and pentagons as the complementary regions

Math. Mag. **86** (2013) 83–96. doi:10.4169/math.mag.86.2.83.

So now you begin to wonder. Are there infinitely many such drawings? Can we construct drawings with however many bigons and pentagons? What if we want to restrict ourselves to some other choices of polygons beside 2-gons and 5-gons, say, 2-gons and n-gons for some n other than 5?

You continue to doodle, until you look up and realize that you are alone in the lecture hall. You are missing the reception with the pink frosted cookies, which was the primary reason you came to the talk. But you continue to doodle. Now you are hooked.

In this paper, we investigate the question of how many combinatorially distinct ways there are to draw a self-intersecting closed curve, such that all of the complementary faces are either 2-gons or n-gons. We only consider curves that self-intersect transversely to create crossings as in FIGURES 1 and 2. We show that n must be odd and that for odd $n > 3$, there are infinitely many such curves. If we fix an odd n, and let p_n be the number of n-gons, we show that p_n must be even, and for $n > 3$, every even value is realized by a curve, with one exception: The case $p_n = 6$ is not realizable for any n.

Perhaps most interesting from our own perspective are the motivations coming out of knot theory. We discuss how each such curve corresponds to a collection of knots, and how questions about the complementary n-gons can be translated into questions about knots. In particular, we prove some surprising results about generating all knots using complementary regions consisting only of triangles, quadrilaterals, and pentagons.

But first, we will consider what Euler's equation has to tell us.

Euler's equation to the rescue

If we draw a closed self-intersecting curve in the plane, we can consider each self-intersection as a vertex, and the two arcs of the curve that cross there to be four edges meeting at that vertex. Hence, we generate a so-called 4-valent planar graph. Since we are interested in the complementary regions of this graph, it will be convenient to think of the graph as living on the sphere rather than on the plane, so that the outer region receives no special treatment. However, for convenience, we will continue to draw the graphs in the plane.

The graph breaks up the surface of the sphere into vertices, edges, and faces, and therefore Euler's famous equation applies. Consider the most general case first. Suppose that we have a finite 4-valent graph embedded in the sphere. So edges are not allowed to cross each other or meet a vertex other than at their endpoints. We assume that no edge has both of its endpoints at the same vertex, but we do allow parallel edges, which is to say two edges that share the same pair of endpoints. Note that we cannot have more than two parallel edges sharing a pair of vertices, since a third would require more than one component to our curve, and for the sake of this discussion, we will not allow ourselves to draw more than one curve at a time.

Faces all have two or more edges. Let p_i be the number of faces with i edges, for $i \geq 2$. If v represents the number of vertices, e the number of edges, and f the number of faces, then Euler's equation tells us that $v - e + f = 2$.

However, 4-valency implies that the number of edges is exactly twice the number of vertices. So we know that $f - e/2 = 2$. The total number of faces is just the sum of the numbers of faces with given edge numbers, so we have

$$f = \sum_{i=2}^{\infty} p_i.$$

The total number of edges is just the number of edges on each type of face, times the number of those faces, but then all divided by 2, since each edge occurs on two faces. So we have

$$e = \frac{1}{2}\sum_{i=2}^{n} ip_i.$$

Plugging these in and cleaning up, we obtain

$$2p_2 + p_3 = 8 + p_5 + 2p_6 + 3p_7 + \cdots. \tag{1}$$

Note several important facts here.

1. The number of quadrilaterals, p_4, does not appear. The number of quadrilaterals is not constrained by the numbers of other faces.
2. At least one of p_2 and p_3 must be nonzero to counterbalance the 8 on the right side of the equation.
3. The number of faces with an odd number of edges, p_{odd}, must be even. This follows immediately from Equation (1) by considering it modulo 2.

Given this much, we could ask a variety of questions. One of the first that comes to mind is the following. Given a particular choice of $(p_2, p_3, p_5, p_6, \ldots, p_n)$ that satisfies equation (1), can it be realized by a 4-valent graph, allowing for varying the number of quadrilaterals? Both graph theorists and geometers have been interested in this question, specifically in relation to convex polytopes, for which the graph appears as the collection of edges. But when considering convex polytopes, parallel edges are precluded, and therefore only the case $p_2 = 0$ is considered. Given this assumption, Grünbaum showed that for any finite collection of integers that satisfies Equation (1), there is a choice of p_4 that allows us to realize a 3-connected (removing two vertices cannot disconnect the graph) 4-valent graph on the sphere with those numbers of complementary regions [**6**]. The Steinitz Theorem (cf. [**12**]) says that these are exactly the conditions necessary to realize the graph as the edges of a convex polytope.

But Grünbaum did not restrict to a single closed curve. He allowed for the overlay of a whole collection of such curves. In [**8**], Jeong considered those graphs that can be drawn as a single closed curve. He showed that for any n, and choice of values for $p_3, p_5, p_6, \ldots, p_n$ that satisfy equation (1), there exists a choice for p_4 and a single closed curve that realizes those values.

But what happens if we do not allow p_4 to take any value? And what happens if we allow p_2 to be nonzero?

DEFINITION. Define a self-intersecting closed curve on the sphere to be an $(a_1, \ldots, a_n)$-*curve*, where $2 \le a_1 < a_2 < \cdots < a_n$ are positive integers, if each complementary region has a_i edges for some a_i.

Note that $(a_1, a_2, \ldots, a_n)$ tells us the types of faces we are allowed, unlike $(p_2, p_3, \ldots, p_m)$, which tells us the number of each type of face. From our consideration of Euler's equation, we have already noted that a_1 must be 2 or 3. Furthermore, as pointed out in [**1**], there must be faces with an odd number of sides.

THEOREM 1. *If $a_1, \ldots, a_n$ are all even, there are no $(a_1, \ldots, a_n)$-curves.*

Proof. This follows by choosing any region with which to begin, and coloring the edges alternately black and gray around the outside of the region as in FIGURE 3. This coloring can be consistently extended across complementary regions to all of the

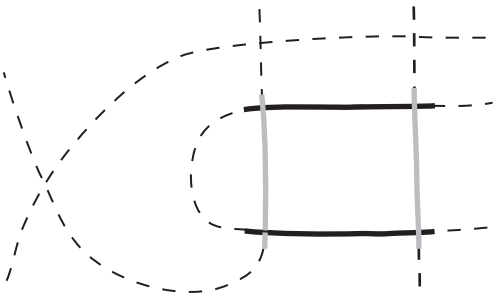


Figure 3 Showing that knots cannot have projections with all even-sided complementary regions

edges in the graph. At each vertex, opposite edges receive the same color. That this is a well-defined coloring follows from the fact we can add regions on one at a time in a manner so that the union of the regions that have had their edges colored remains a topological disk at all times. The result is that there are at least two curves making up the diagram, one black and one gray. ■

Note that a similar proof demonstrates that if $a_1, a_2, \ldots, a_n$ have a nontrivial common factor d, then there are at least d curves in the curve diagram. But now is a good time to turn to applications in knot theory. We will return to the question of generating $(a_1, a_2, \ldots, a_n)$-curves after that.

Knot theory and doodles

A knot is a simple closed curve embedded in 3-space, up to ambient isotopy. Equivalently, two knots are considered equivalent if one can be deformed to look like the other without passing it through itself. One of the most common ways to understand knots is through their projections, where we project down to a plane, and see a self-intersecting curve. We typically consider only the regular projections, where there are no points of tangency, and where there are at worst a finite number of double points, called crossings. We keep track of which strand is over and which is under at each crossing.

Clearly, each of the self-intersecting curves we are drawing generates a knot projection by making a choice of which way each crossing goes at the self-intersections. So a single curve with n self-intersections generates 2^n knot projections, as in FIGURE 4. Note, however, that many of these different projections represent the same knot up to knot equivalence.

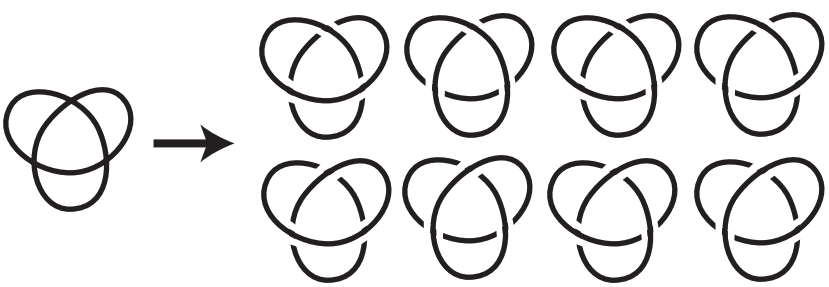


Figure 4 Turning a curve into knot projections

Thus, any information obtained about the kinds of curves we can generate can be translated into information about knots. For instance, the fact that all complementary regions having an even number of sides implies more than one curve, demonstrates that no knot can have a projection with all even-sided complementary regions.

Although we cannot rid ourselves of all odd-sided faces, we can get close with the following result, the proof of which we do not include here.

THEOREM 2. *Every knot has a projection with exactly two odd-sided regions, and they can be chosen to be triangles* [**1**].

But what about the other extreme?

THEOREM 3. *Every knot has a projection with all odd-sided regions* [**1**].

Proof. The argument is very simple. Start with any projection. For any region that has an even number of edges, say n, apply the move in FIGURE 5 to obtain a new projection of the same knot such that the parity of the edge numbers for all the other original regions remains unchanged, all of the new regions have an odd number of edges, and the number of edges for this region becomes odd. ∎

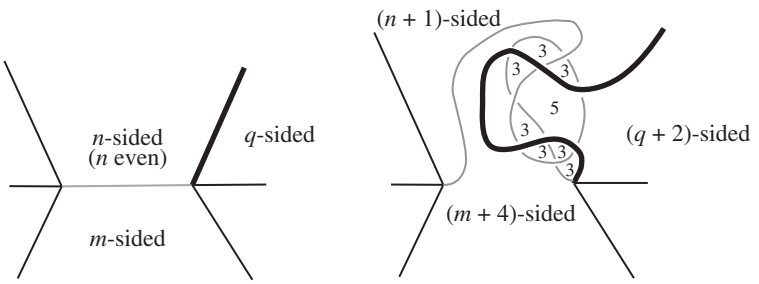


Figure 5 Making an even-sided region into an odd-sided region

DEFINITION. A knot is said to be an $(a_1, a_2, \ldots, a_n)$-*knot* if it possesses a projection that is an $(a_1, a_2, \ldots, a_n)$-curve.

Which knots are encompassed by a particular finite sequence $(a_1, \ldots, a_n)$? Could every knot be an $(a_1, \ldots, a_n)$-knot for some choice of $(a_1, \ldots, a_n)$?

DEFINITION. A sequence $(a_1, \ldots, a_n)$ is said to be *universal for knots* if every knot is an $(a_1, \ldots, a_n)$-knot.

Could (3, 4, 5) be universal? In other words, does every knot have a diagram that can be built with only triangles, quadrilaterals, and pentagons for its complementary regions? In fact, surprisingly enough, the answer is yes.

THEOREM 4. *The sequence* (3, 4, 5) *is universal for knots* [**1**].

Proof. The first step is to realize that there are infinitely many (3, 4)-curves. In FIGURE 6, we see a curve that has exactly eight triangles and a whole lot of quadrilaterals. By spiraling the curve around further, we can increase the number of quadrilaterals, while keeping the number of triangles at 8, as equation (1) says we must.

The second fact we need is a result from knot theory, which says that by changing crossings, we can turn any projection into a projection of the trivial knot. Here is a simple way to convince ourselves that we can do so. Suppose that we have a picture in the plane of a closed curve with self-intersections. Take a piece of string and, starting

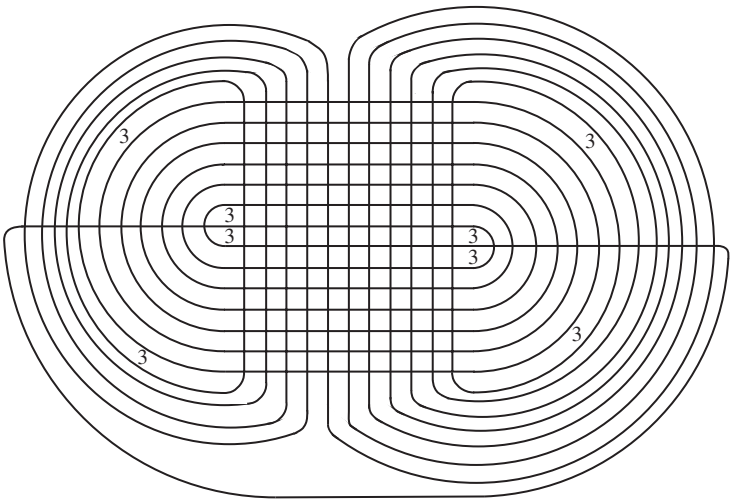


Figure 6 A (3, 4)-curve in the plane

at any point on the curve, lay the string down along the curve path. Each new crossing that is created will have the new strand of string that is laid down as the overstrand. Once the entire curve is covered, trim off any excess string and tape the two ends of the string together. Clearly, the knot given by the string has our curve as a projection. So it can be obtained from that projection by setting the crossings of the projection to match its actual crossings. But if we lift the string back into space, reversing the order in which we laid it down, each crossing is eliminated one at a time, with no resulting tangling, since in the creation of each crossing, the new strand was always the overstrand. Hence, the resulting knot is the trivial knot. So every projection can be made into a trivial knot by setting the crossings appropriately.

Now we are ready to prove that every knot has a (3, 4, 5)-diagram. Start with any knot K and take any projection P of it. We first rectangularize the projection so it lies in the unit grid, and all of the edges of the projection are parallel to either the x-axis or y-axis, as in FIGURE 7. That we can do this follows from the fact that we can approximate the original projection with a rectangular projection that stays in an arbitrarily small neighborhood of the original.

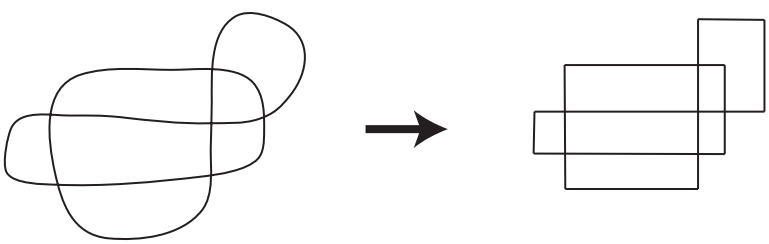


Figure 7 Straightening a projection

Now we choose a (3, 4)-curve Q that has a grid of quadrilaterals big enough to contain the entire projection P. We choose the crossings on Q to trivialize the projection. We lay P down on Q offset by $1/2$ in both the x and y directions, so that the crossings and corners of P occur at the center of complementary squares of Q and so that at all new crossings between P and Q, P crosses over Q. As in FIGURE 8, where the

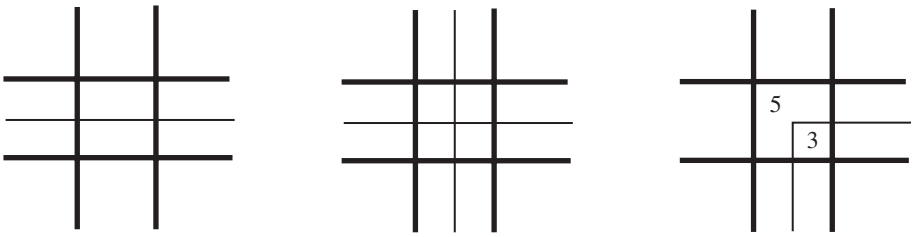


Figure 8 When P overlays Q, we obtain a $(3, 4, 5)$-projection.

thin lines correspond to P and the thick lines to Q, when P passes straight through a square of Q, it cuts it into two quadrilaterals. A crossing of P cuts a square of Q into four squares. A corner of P cuts a square of Q into a triangle and a pentagon. Hence, we now have a $(3, 4, 5)$-projection of a 2-component link, one component of which is K and the other component of which is the trivial knot.

We now compose these two knots. Choose a square of Q that is divided in half when P crosses through it. Then, as in FIGURE 9, we add one crossing and connect P to Q. This divides the two squares into two pentagons and two triangles. The resulting knot is the composition of K with the trivial knot, which is in fact K back again. So we have obtained a projection of K that is a $(3, 4, 5)$-projection. ■

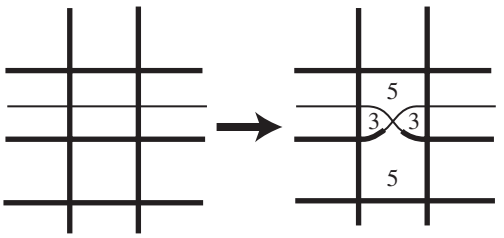


Figure 9 Connecting P to Q

In [**1**], it was further proved that for any $n \geq 5$, $(3, 4, n)$ is universal for knots. This is somewhat surprising. It says, for instance, that any knot has a projection such that all of the complementary regions are triangles, quadrilaterals, and 1,000,001-gons. It was further proved that $(2, 4, 5)$ is also universal for knots.

We cannot help asking whether we can do even better. Specifically, do there exist two-integer sequences (a_1, a_2) that are universal for knots? The Euler equation and our previous results limit the possible sequences to $(2, n)$, for odd $n \geq 5$ and $(3, n)$, where $n \neq 0 \pmod 3$ and $n \geq 4$.

But notice that if a sequence such as $(2, 5)$ is going to be universal for knots, then, at the very least, we need an infinite number of $(2, 5)$-curves. A finite number of such curves can only generate a finite number of knots. Although ultimately we will not determine whether $(2, n)$ or $(3, n)$ are universal for knots—that question remains open—it motivates our interest in whether there are infinitely many such curves. So that is what we turn to next.

Finding $(2, n)$-curves

Our goal is now to decide whether there are infinitely many combinatorially distinct $(2, n)$-curves. Let's start with $(2, 3)$-curves.

(2, 3)-curves In the case of $(2, 3)$-curves, the answer is clearly no. Equation (1) becomes

$$2p_2 + p_3 = 8.$$

The only possibilities are $(p_2, p_3) = (3, 2), (2, 4), (1, 6)$. (If we allow either p_2 or p_3 to be 0, then we obtain the edge graph of the octahedron, which yields a projection of the Borromean rings; or we obtain two linked rings, so neither is a single curve.) The cases $(3, 2)$ and $(2, 4)$ are realized by projections of the trefoil knot and the figure-eight knot appearing in FIGURE 10. It is easy to eliminate $(1, 6)$, since if we start with a bigon and glue triangles to each of its edges, then we can only glue on two more triangles before the curve closes up, and the outer region is another bigon.

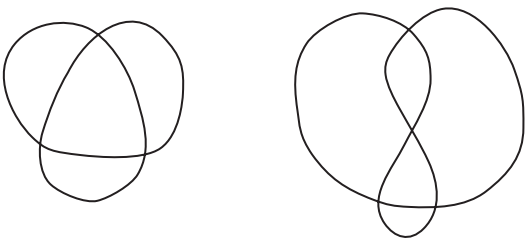


Figure 10 Realizing $(p_2, p_3) = (3, 2), (2, 4)$

(2, 5)-curves We now consider the first interesting case, which is given by $(2, 5)$-curves. Since only p_2 and p_5 are nonzero, equation (1) becomes

$$2p_2 = 8 + p_5.$$

This immediately implies that p_5 must be even, and that once we determine p_5, the value of p_2 is also determined. So which even values of p_5 can be realized?

THEOREM 5. *There exist* $(2, 5)$*-curves for all positive even* $p_5 \neq 6$ [**11**].

Proof. For $p_5 = 2$ and 4, consider FIGURE 11, which when made alternating, are projections of the 5_1 and 8_3 knots.

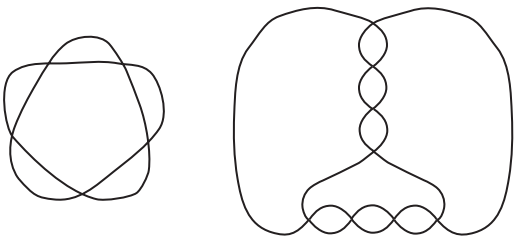


Figure 11 $(2, 5)$-curves for $p_5 = 2, 4$

So we need only consider $p_5 \geq 6$. The first construction we use is an infinite class of $(2, 5)$-curves that resemble the diagram depicted in FIGURE 12. We can iterate the pattern to get arbitrarily complicated curves. We can show by an inductive argument that a diagram constructed in this manner is a single curve if and only if p_5 is not a multiple of 6.

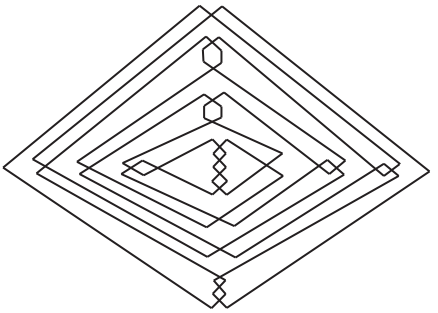
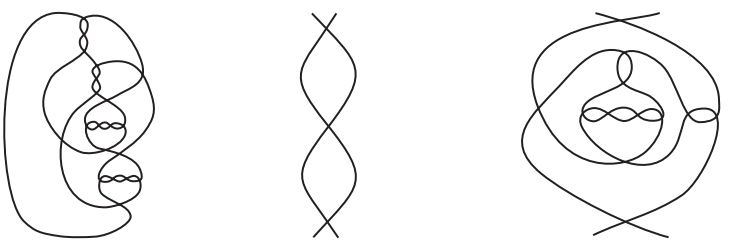


Figure 12 There are infinitely many (2, 5)-curves.

To show that there are single-component diagrams for $p_5 = 0 \pmod 6$ when $p_5 > 6$, we consider the diagram in FIGURE 13(a). This is a single-component (2, 5)-diagram with $p_5 = 12$. Now consider a subsection of this diagram that appears as in FIGURE 13(b). By replacing this subsection with one that looks like FIGURE 13(c), we can increase p_5 by 6, without altering the number of components.

We can repeat this replacement arbitrarily many times to get a single-component (2, 5)-diagram for any positive $p_5 \equiv 0 \pmod 6$, $p_5 \neq 6$. So between these two constructions, there exist single-component diagrams for any even $p_5 > 6$.



(a) A base diagram (b) A subdiagram (c) A subdiagram extension

Figure 13 Extending a diagram

We reserve the proof that $p_5 = 6$ cannot be realized by a (2, 5)-curve for Theorem 7 appearing below. ■

(2, *n*)-curves We now consider (2, n)-curves for n odd, $n > 5$. We prove the following:

THEOREM 6. *For n odd, $n \geq 5$, there exists a (2, n)-curve for every even p_n except $p_n = 6$* [**11**].

Proof. The base and extension diagrams for (2, 5)-curves can be turned into (2, n)-diagrams by extending certain crossings with chains of 2-gons of length $n - 5$. In FIGURE 14, some additional base and extension diagrams are shown, with choices for which crossings to extend, where the gray line segments run perpendicular to the orientation of the inserted chain. ■

Proving there are no (2, *n*)-curves with six *n*-gons

The last step is to show that the case of six n-gons never occurs for (2, n)-curves. It will take some work to prove this, but since this case is so anomalous, and since this

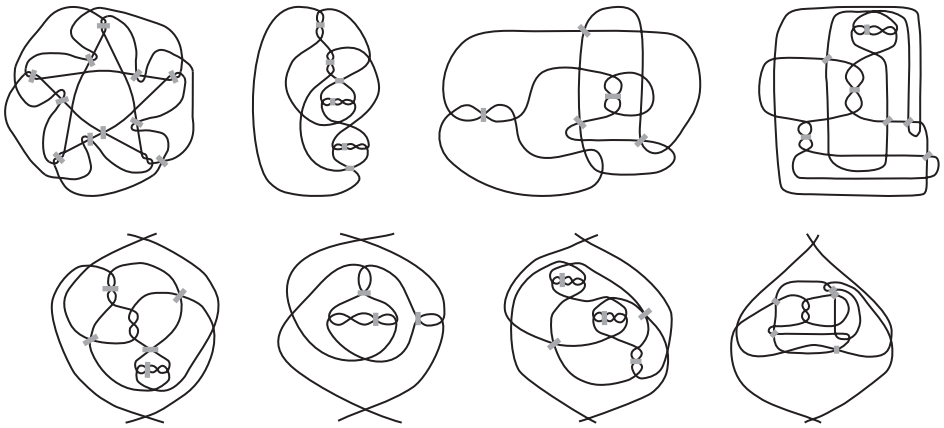


Figure 14 Replace marked crossings with $n-5$ 2-gons to produce base and insertion diagrams for $(2, n)$.

result has not to our knowledge previously appeared in the literature, we include it here.

THEOREM 7. *There does not exist a* $(2, n)$*-curve with* $p_n = 6$ *for any odd integer* $n \geq 3$.

Proof. We eliminated the case $n = 3$ above, so we will assume that n is odd and $n \geq 5$. Suppose there is a $(2, n)$-curve γ such that $p_n = 6$. Let G be the 4-valent spherical graph obtained from γ. Collapse the two parallel edges of each bigon to a single edge, and then replace chains of edges connected by 2-valent vertices by single edges to obtain a graph G' that has only vertices of valency 3 and 4, and six faces for its complementary regions, as in FIGURE 15. From now on, the faces corresponding to a graph are the closures of the complementary faces. If we wish to reconstruct G from G', then exactly one of the three edges coming out of each valency three vertex is replaced by a chain of bigons. We call such an edge of G' a *bigonal* edge. See FIGURE 15.

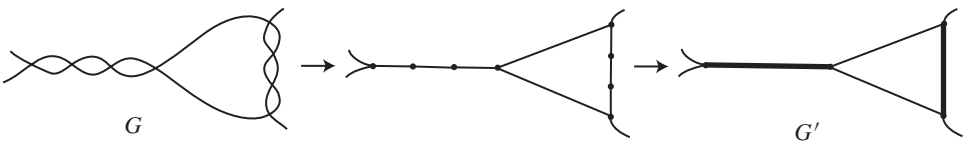


Figure 15 Going from G to G' with thickened edges bigonal

Showing that pairs of faces intersect in a connected set First, suppose that there exist two faces A' and B' of G' that intersect in a disconnected set. Then the complement of their union is a finite collection of disconnected sets on the sphere. Suppose that one of those sets consists of the interior of a single face C. Then, since only two edges of G' lie on the boundary of C, and C came from an n-gon corresponding to G, one of these two edges of G' must be a bigonal edge that came from a chain of bigons of G of length $n-1$. Hence, both vertices have valency three. However, this implies that A' and B' must each share at least two edges with one another, one intersecting the first vertex and one intersecting the second. But then one of A' and B' came from a face of G that had more than n edges, a contradiction.

Thus, each complementary region of $A' \cup B'$ must contain the interiors of at least two faces. Since there are only six such faces total, there must be exactly two complementary regions of $A' \cup B'$, each containing the interiors of two such faces. Hence we must have a pattern as in the left side of FIGURE 16, where all bold edges are either bigonal edges and a, b, c, and d represent the number of bigons in each labeled edge, or a bold edge corresponds to a single vertex and the corresponding label is 0. Note then that $a + b = n - 4$, and then $c = d = n - 2$. However, since n is odd, we see that G cannot have come from a single curve, but must have come from two distinct curves, as in the right side of FIGURE 16.

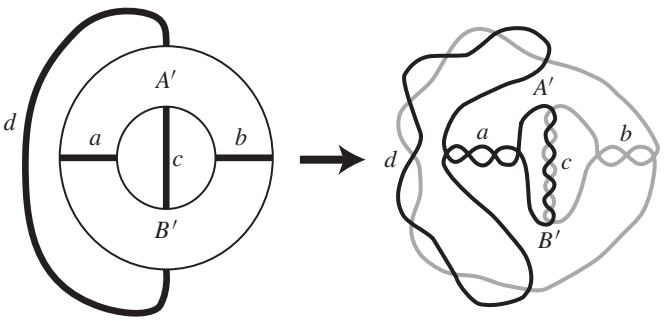


Figure 16 Complementary faces of G' intersecting in a disconnected set yield a link rather than a knot.

Hence, we can assume that any two faces corresponding to G' intersect in a connected set. Thus, two faces either do not intersect, they intersect in a single vertex, or they intersect in a single edge and the vertices at its ends. In particular, this means that all of the faces of G' are triangles, quadrilaterals, or pentagons, since two faces can intersect in at most an edge and there are only six faces total.

Showing that every face has a bigonal edge We next show that every face must have a bigonal edge on its boundary. For if there is a face without a bigonal edge, call it A, then since $n \geq 5$, the other five faces must each share a non-bigonal edge with A and the only possibility for n is 5. Then choosing A to be the outer face in the plane, we obtain a tiling of the pentagon by these five faces. Each edge in the interior of the pentagon that hits one of the vertices of the pentagon must be a bigonal edge, since otherwise one of the interior faces would intersect the exterior face in a disconnected set. Since two bigonal edges cannot share a vertex, none of the interior endpoints of these edges can coincide. Thus, each of these faces must be a quadrilateral or a pentagon and the total number of vertices must be at least ten. Suppose that there is an additional vertex. Then three or four of the five faces are incident with this vertex. But then two adjacent faces, which already intersect in a bigonal edge, also intersect in this additional vertex, a contradiction. Hence there can be no other vertices. By the Euler characteristic of the pentagon, we have 10 vertices, each of valency 3, so 15 edges and 5 faces, so $v - e + f = 1$ yields $0 = 1$, a contradiction.

The grand finale to the proof Let v_3 and v_4 be the number of valency 3 and 4 vertices, respectively. The Euler characteristic implies $v_3 + 2v_4 = 8$. Each face of G' must have at least one bigonal edge, and since the endpoint of a bigonal edge must be a valency 3 vertex that hits no other bigonal edge, there must be at least six valency 3 vertices. Hence, the only possibilities are $(v_3, v_4) = (6, 1)$ and $(8, 0)$. In the case $(v_3, v_4) = (6, 1)$, the six valency 3 vertices break up into pairs shared by the bigonal

edges. Form a graph G'' from G' by removing the bigonal edges, and their vertices, resulting in three faces. Then G'' has one vertex v, which is valency 4. So G'' must just consist of two looped edges ending at the vertex v that cut the sphere into three faces, each of which must contain one of the bigonal edges. But then the region bounded by both loops, which appears in FIGURE 17 as the outer face, once split by its bigonal edge, will generate two faces A' and B' of G' that intersect in a disconnected set, as in FIGURE 17, a contradiction.

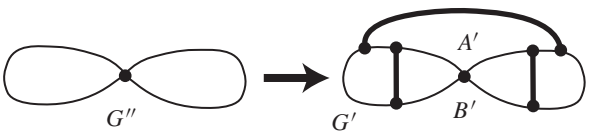


Figure 17 Six valency three vertices cannot occur.

The final case to consider is $(v_3, v_4) = (8, 0)$. This implies that there are four bigonal edges. Again, eliminating the bigonal edges and their vertices results in a collection of curves with no vertices. There can be at most two such curves, or else a complementary region of the curves contains no bigonal edge or G' is not connected, both of which cause a contradiction. In the case of two such curves, they cut the sphere into three regions, one of which is an annulus. Again, each region must contain a bigonal edge. Since G' is connected and the closure of any face in G' must be embedded, there must be two bigonal edges crossing from one boundary of the annulus to the other. However, this creates a pair of faces for G' that intersect in a disconnected set.

In the case of one such curve, we can have either two bigonal edges to each side of the curve, or one bigonal edge to one side and three to the other side, as occurs in FIGURE 18. But in either case, no matter how many bigons are placed on the bigonal edges, the n-gon faces of G cannot be made to have the same number of edges. ■

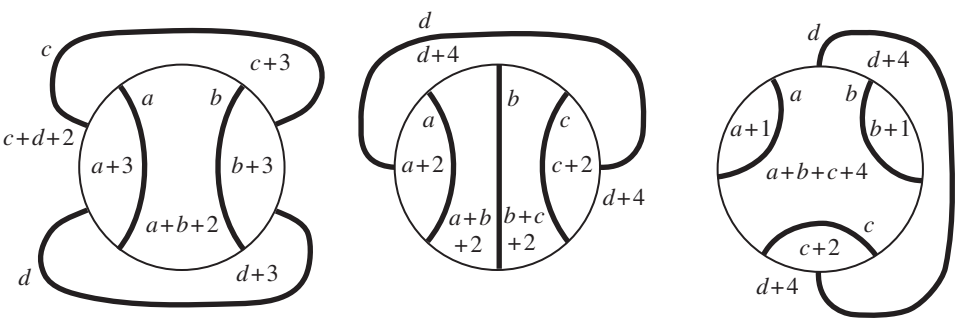


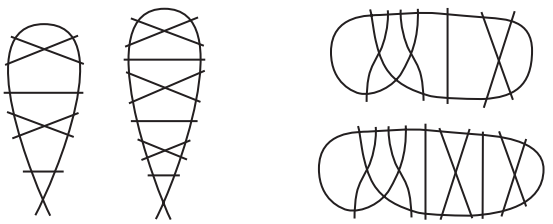
Figure 18 The number of edges on all faces cannot be made to match.

Conclusions

In addition to the results above on $(2, n)$-curves, Noël MacNaughton proved the following about $(3, n)$-curves, keeping in mind that such a curve must have $n \not\equiv 0 \pmod 3$.

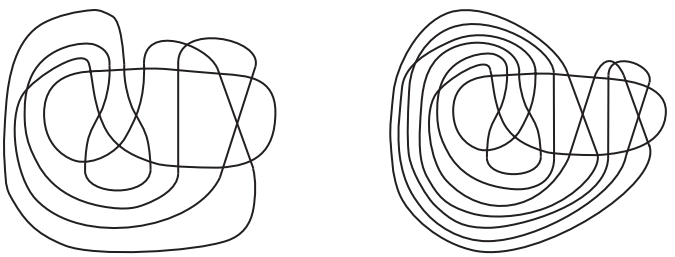
THEOREM 8. *There exist infinitely many* $(3, n)$*-curves for all* $n \not\equiv 0 \pmod 3$.

The basic idea for $(3, 5)$-curves is to use snowshoe patterns as in FIGURE 19. By repeating the pattern inside the loop, and finding a similar pattern for outside the loop (which we will turn inside out from the pattern shown), we get partial diagrams that look like those in FIGURE 19. By matching up these diagrams, we get single-component diagrams like those in FIGURE 20.



(a) (b)

Figure 19 Repeating the pattern inside and outside the loop



(a) (b)

Figure 20 Diagrams obtained by pairing inner and outer loop subdiagrams

We can complicate these diagrams to produce $(3, n)$-curves. Consider a subdiagram as in FIGURE 21(a). By replacing it with a subdiagram as in FIGURE 21(b), we add three sides to the faces in the top left and bottom right corners. Using arguments such as these, we can show that there are infinitely many $(3, n)$-curves for all $n \not\equiv 0 \pmod 3$.

(a) (b)

Figure 21

Hence, both $(2, n)$-curves and $(3, n)$-curves remain candidates for being universal for knots. But it remains open as to whether they are.

Simple questions about the combinatorial nature of immersed curves in the plane have been and continue to be fertile ground for research. In addition to the references mentioned above, see [**2, 5, 6, 9, 13**]. Many questions remain open. Let k be the number of curves making up a curve diagram. For a given sequence $(p_2, \ldots, p_n)$ that satisfies equation (1), one can ask for exactly which k that sequence is realizable. We can also consider all of these questions for curve systems on surfaces other than the sphere (cf. [**10, 14**]).

From the point of view of knot and link theory, see also [**4**]. We can further ask whether $(2, n)$ and $(3, n)$ are universal for knots for various n. We can ask whether (a, b, c) is universal when $b \neq 4$. Lots of interesting questions remain. So time to quit doodling and get to work!

Acknowledgment This paper is based on results from [**1**], work by Charmaine Sia in the NSF funded SMALL Undergraduate Research Program at Williams College in summer, 2009, and on results from the Williams College undergraduate thesis of Nöel MacNaughton, 2010.

REFERENCES

1. C. Adams, R. Shinjo, and K. Tanaka, Complementary regions of knot and link diagrams, *Annals of Combinatorics* **15** (2011) 549–563. http://dx.doi.org/10.1007/s00026-011-0109-2
2. M. Deza, M. Dutour, and M. Shtogrin, 4-valent plane graphs with 2-, 3- and 4-gonal faces, *Advances in Algebra,* World Scientific, River Edge, NJ, 2003, 73–97.
3. V. Eberhard, *Zur morphologie der polyeder,* Leipzig, 1891.
4. S. Eliahou, F. Harary, and L. Kauffman, Lune-free knot graphs, *J. Knot Theory Ramifications* **17** (2008) 55–74. http://dx.doi.org/10.1142/S0218216508005963
5. T. C. Enns, 4-valent graphs, *J. Graph Theory* **6** (1982) 255–281. http://dx.doi.org/10.1002/jgt.3190060303
6. B. Grünbaum, Planar maps with prescribed types of vertices and faces, *Mathematika* **16** (1969) 28–36. http://dx.doi.org/10.1112/S0025579300004587
7. ———, *Convex Polytopes,* Graduate Texts in Mathematics Vol. 221, Springer-Verlag, New York, 2003.
8. D. Jeong, Realizations with a cut-through Eulerian circuit, *Discrete Math.* **137** (1995) 265–275. http://dx.doi.org/10.1016/0012-365X(93)E0129-R
9. E. Jucovic, On the face vector of a 4-valent 3-polytope, *Studia Sci. Math. Hungar.* **8** (1973) 53–57.
10. E. Jucovic and M. Trenkler, On 4-valent graphs imbedded in orientable 2-manifolds, *Studia Sci. Math. Hungar.* **7** (1972) 225–232.
11. N. MacNaughton, *From Doodles to Diagrams*, Williams College Undergraduate Mathematics Thesis, 2010.
12. E. Steinitz, Polyeder und Raumeinteilungen, *Enzykl. math. Wiss.* **3** (Geometrie) (1922) 1–139.
13. M. Trenkler, Convex 4-valent polytopes with prescribed types of faces, *Comment. Math. Univ. Carolin.* **25** (1984) 171–179.
14. J. Zaks, The analogue of Eberhard's theorem for 4-valent graphs on the torus, *Israel J. Math.* **9** (1971) 299–305. http://dx.doi.org/10.1007/BF02771680

Summary What closed curves can be drawn in the plane such that they cut the plane into complementary regions that are n-gons, including the outer region, where n is allowed to take some finite number of values? A curve is an $(a_1, a_2, \ldots, a_n)$-curve if the number of edges for its complementary regions all lie in $a_1, a_2, \ldots, a_n$. We show that there are infinitely many curves for $(2, n)$, where n is any odd integer greater than 3, and for $(3, n)$, for any $n > 3$ relatively prime to 3. We also consider the implications for knot theory, showing that every knot has a $(3, 4, 5)$-diagram. We ask what values of $(a_1, a_2, \ldots, a_n)$ will generate diagrams for every knot.

COLIN ADAMS is the Thomas T. Read Professor of Mathematics at Williams College. His research focuses primarily on knots and hyperbolic 3-manifolds. He is the author or co-author of the books *Riot at the Calc Exam*, *Introduction to Topology: Pure and Applied*, *How to Ace Calculus*, *How to Ace the Rest of Calculus*, *The Knot Book*, and the comic book *Why Knot?*. He is also the math humor columnist for the Mathematical Intelligencer.

NÖEL MacNAUGHTON received her undergraduate degree in mathematics from Williams College. For her undergraduate thesis, she worked on the question of how common are the $(2, n)$ and $(3, n)$-curves discussed in this paper. She currently works for Bank of America in New York City, using the abilities she honed as a math major on a daily basis.

CHARMAINE SIA is a graduate student in mathematics at Harvard University, and was an undergraduate at MIT at the time this research was completed. Her research area is algebraic topology, but she is also interested in topics related to knot theory. Her current knot-related interest is the analogy between knots and primes in arithmetic topology.

Deranged Socks

SALLY COCKBURN
Hamilton College
Clinton, NY 13323
scockbur@hamilton.edu

JOSHUA LESPERANCE
Oberlin College
Oberlin, OH 44074
joshua.p.lesperance@gmail.com

Gloves and derangements

In the spring semester of 2005, the first author presented the following problem (and solution) as an example for her graph theory and combinatorics class, as found in Alan Tucker's *Applied Combinatorics* [**5**]:

GLOVE PROBLEM. Given five pairs of gloves, how many ways are there for five people each to choose two gloves with no one getting a matching pair?

In this problem, we assume that we can distinguish between left and right gloves. If we require that each person choose one left glove and one right glove, then the answer is 5,280; if we allow for the possibility of a person choosing two gloves for the same hand, then it shoots up to 65,280. Obtaining these answers is a nice illustration of combinatorial techniques.

Mismatched gloves may remind us of derangements. A *derangement* of n objects is a permutation in which every object gets moved. To count the number of derangements of n distinct objects, we turn to the principle of inclusion-exclusion, used for counting the number of elements in some universal set $\mathcal{U}$, which satisfy none of n different properties. We let A_i denote the elements of $\mathcal{U}$ satisfying property i, and S_k the sum of the cardinalities of all k-fold intersections of the A_i; that is,

$$S_k = \sum_{i_1 < \cdots < i_k} |A_{i_1} \cap \cdots \cap A_{i_k}|.$$

For example,

$$S_1 = |A_1| + |A_2| + \cdots + |A_n|$$

and

$$S_2 = |A_1 \cap A_2| + |A_1 \cap A_3| + \cdots + |A_{n-1} \cap A_n|$$

(including all $\binom{n}{2}$ pairs). Then the principle states that

$$\left|\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}\right| = |\mathcal{U}| - S_1 + S_2 - S_3 + \cdots + (-1)^n S_n. \tag{1}$$

The left side of (1) is the number of elements of $\mathcal{U}$ that are not contained in any of the sets A_i.

Math. Mag. **86** (2013) 97–109. doi:10.4169/math.mag.86.2.97.

To count derangements, we let $\mathcal{U}_n$ denote the set of all permutations of $\{1, \ldots, n\}$, and A_i denote the subset of these permutations in which object i is in its original position. Then $|\mathcal{U}_n| = n!$ and $|A_i| = (n-1)!$ for all $1 \le i \le n$. More generally, $|A_{i_1} \cap \cdots \cap A_{i_k}| = (n-k)!$ for all subsets $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$, and since there are $\binom{n}{k} = n!/[k!\,(n-k)!]$ such subsets, we get $S_k = \binom{n}{k}(n-k)! = n!/k!$. Hence the nth derangement number is

$$D_n = n!\left[\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\right].$$

The first few values are $D_1 = 0$, $D_2 = 1$, $D_3 = 2$, $D_4 = 9$, $D_5 = 44$. An immediate consequence is that the fraction of all permutations that are derangements quickly converges to $1/e$ as $n \to \infty$.

Derangements can be used to solve the version of the Glove Problem in which n people each choose one left glove and one right glove. We begin by labeling the gloves $1_L, 1_R, \ldots, n_L, n_R$. To create n mismatched pairs of gloves, we first line up all of the left-hand gloves in their natural order: $1_L, 2_L, \ldots, n_L$. We then pair them up with a derangement of the right-hand gloves, and there are D_n ways to do this. Finally, there are $n!$ ways for us to distribute all n pairs of gloves, since no two glove-pairs will be the same, giving us the solution

$$n!D_n = [n!]^2\left[\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\right].$$

When $n = 5$, this formula gives $(5!)D_5 = 5{,}280$, one answer to the original Glove Problem.

If we allow people to choose gloves regardless of handedness, then the answer is more subtle, but we can again invoke the principle of inclusion-exclusion. Let $\mathcal{U}_n$ be the set of all possible ways of distributing 2 gloves to each of n people, from a set of $2n$ distinct gloves. To compute $|\mathcal{U}_n|$, we line up the gloves in order $1_L, 1_R, \ldots, n_L, n_R$. We then assign to this glove lineup a permutation of the multiset consisting of two copies of each person's name, which gives

$$|\mathcal{U}_n| = (2n)!/2^n.$$

If we let A_i denote the subset of glove distributions in which somebody gets matching pair i, then to compute S_k, we first choose which k gloves are matched, distribute these matching pairs to k lucky people, then distribute the remaining $n-k$ pairs in any fashion. If we use $\binom{n}{k}$ to denote the number of k-combinations of n objects and $P(n,k) = n!/(n-k)!$ to denote the number of k-permutations of n objects, then we get as our answer

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}P(n,k)\frac{[2(n-k)]!}{2^{n-k}} = \sum_{k=0}^{n}(-1)^k\binom{n}{k}P(n,k)|\mathcal{U}_{n-k}|.$$

When $n = 5$, this gives a value of 65,280, which is the other answer to the original Glove Problem.

Socks and derangements

Suppose that we remove the crucial assumption that right- and left-hand gloves are distinguishable. As suggested by one of our students, this happens naturally if we

phrase the problem in terms of socks, rather than gloves. The resulting problem is surprisingly more difficult.

DERANGED SOCK PROBLEM. Given n distinct pairs of socks, how many ways are there for n people each to choose two socks with no one getting a matching pair? For the remainder of this paper, the solution to this problem shall be denoted d_n.

Note that there is only one version of this problem: If left and right socks are indistinguishable, then it doesn't make sense to consider the version where each person ends up with one left sock and one right sock. If we attack this problem with the principle of inclusion-exclusion, then by analogy to the glove problem, we can start with what might seem to be an easier problem.

SOCK DISTRIBUTION PROBLEM. Given n distinct pairs of socks, how many ways are there for n people each to choose two socks? The solution to this problem shall be denoted u_n.

Inclusion-exclusion allows us to compute d_n as a function of u_j for $1 \le j \le n$, but we can also compute u_n as a function of d_j by dividing into mutually exclusive, exhaustive cases according to how many people get a matching pair. This leads to the pair of formulas:

$$d_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} P(n,k) u_{n-k}; \tag{2}$$

$$u_n = \sum_{k=0}^{n} \binom{n}{k} P(n,k) d_{n-k}. \tag{3}$$

Thus, if we can compute d_n for all n or u_n for all n, then we can compute the other sequence. It seems a good bet to try for u_n first. However, sock distribution is not as easy as glove distribution. To illustrate the difficulty, consider the case $n = 4$ and suppose that the people are named A, B, C, and D. We begin by trying to mimic the approach we used for gloves above: line up the socks in canonical order, minus the L and R labels, and then assign to that list an arrangement of two copies of each name. We quickly see that this fails, for

$$\begin{array}{cccccccc} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ \hline A & B & A & B & C & C & D & D \end{array} \quad \text{and} \quad \begin{array}{cccccccc} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ \hline B & A & A & B & C & C & D & D \end{array}$$

result in the same people having the same socks, as will the name permutations $ABBACCDD$ and $BABACCDD$. Similarly, there are 8 name permutations that result in the same sock distribution as $ABACBCDD$, and 16 for $ABCDABCD$. Surely there is a simpler way.

As a stab at computing d_n directly, we break it down by first forming sock pairs and then distributing them, as we did in the first version of the glove problem. (Here and in the rest of this paper, a *sock pair* refers to any two socks to be distributed to a single person, whether matching or not.) Line up one sock of each type in order, then pair them up with a derangement of the remaining socks. For example,

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \hline 4 & 1 & 2 & 3 \end{array} \quad \text{or} \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \hline 2 & 3 & 4 & 1 \end{array} \quad \text{or} \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \hline 4 & 3 & 2 & 1 \end{array}.$$

Even in this small case, several problems become apparent. The first two derangements, which we hoped would be distinct, both result in the same four sock pairs:

$\{1, 4\}$, $\{1, 2\}$, $\{2, 3\}$, and $\{3, 4\}$. Furthermore, the third derangement results in repetition of sock pairs: $\{1, 4\}$, $\{1, 4\}$, $\{2, 3\}$, and $\{2, 3\}$. Thus, while there are 4! ways to distribute the sock pairs of the first two derangements, there are only $4!/2^2$ ways to distribute the sock pairs from the third derangement. As the number of socks and people increase, these problems only get worse.

These cursory attempts demonstrate that switching from gloves to socks significantly complicates matters. In fact, this problem has been addressed before, as one case of a larger problem. In the most general form of the problem, we define $H(n, r)$ to be the number of ways of distributing r copies of n distinct objects among n people, each person getting exactly r objects in total. The first correct solution to the $r = 2$ version of the problem was given by Kenji Mano in 1961 (see [**2**]), in the form of a pair of rather complicated recursive formulas. In this paper, we show how the problem can be attacked using a variety of basic weapons in the discrete mathematics arsenal. We make use of partitions, cyclic permutations, recurrence relations, and generating functions. We obtain solutions in the form of both recursive and (two different) nonrecursive formulas, closed forms of an exotic variation on the generating functions, and finally, with the help of a couple of big guns from complex analysis, asymptotic formulas.

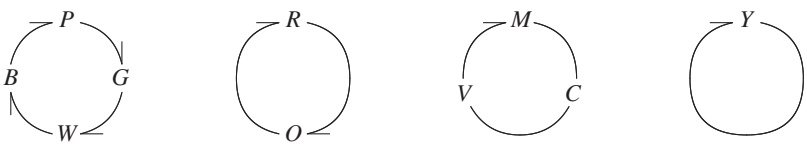
Partitions and cyclic permutations

In this section we return to our earlier approach for computing both d_n and u_n. We will first form n sock pairs and then distribute them to n people. We will find that both parts offer challenges that we did not encounter when sorting gloves. We begin with an example.

Consider the following pairing of socks, using the colors Red, Blue, Green, Yellow, Orange, Pink, White, Violet, Magenta, and Carbon:

$$\{\{R,O\}, \{M,C\}, \{P,G\}, \{P,B\}, \{R,O\}, \{W,B\}, \{C,V\}, \{M,V\}, \{Y,Y\}, \{G,W\}\}.$$

We see that every color appears exactly twice. Additionally, we have a matched pair, which is allowable for u_n but not for d_n. Chains, or *cycles*, of colors can naturally be formed from this set as follows: We pick any color to begin with, say Pink, and find a color that is paired with Pink. There is a pair $\{P,G\}$, so we might choose Green. Then we find another color that is paired with Green, and so on, using each pair only once. In this case, Pink is paired with Green, Green is paired with White, White is paired with Blue, and Blue is paired with Pink, ending the first cycle. After a cycle ends, we pick a color that we haven't used yet and repeat, until we have used all of the colors. Red is paired with orange, which is paired with Red again. Magenta is paired with Carbon, Carbon is paired with Violet, and Violet is paired with Magenta. Finally, Yellow is paired only with Yellow.



These are cycles, not just sets.

Since every color appears exactly twice in this set of sock pairs and the number of colors is finite, this process will always produce cycles (beginning and ending with

the same color). Furthermore, a cycle accounts for both copies of each color, each cycle can be thought of as a permutation of the colors involved, and no two cycles have any colors in common. Therefore, the entire set of sock pairs naturally corresponds to a permutation of the color set, written as a product of disjoint cycles: $(PGWB)(RO)(MCV)(Y)$. In fact, this set of sock pairs corresponds to more than one permutation, since each cycle can be reversed. (Order in the set of sock pairs, as well as in each individual sock pair, does not matter.)

Conversely, any permutation of the color set, which we know can be written as a product of disjoint cycles, naturally corresponds to a set of sock pairs, by placing adjacent colors into a pair. For example, the permutation

$$(RYV)(GMO)(CP)(B)(W)$$

gives us the set of sock pairs

$$\{\{R,Y\},\{Y,V\},\{V,R\},\{G,M\},\{M,O\},\{O,G\},\{C,P\},\{P,C\},\{B,B\},\{W,W\}\}$$

This correspondence between the set of sock pairs and the set of permutations of the color set holds for all such examples, and while it is not one-to-one, we can still use it to reach our goal. But first, we simplify our notation, so that we can more easily address the general problem.

We label the pairs of socks from 1 to n (that is, we have two socks labeled 1, two socks labeled 2, etc), referring to these as the *colors* of the socks. Consider any possible pairing of the $2n$ socks, $M = \{\{i_1, i_2\}, \{i_3, i_4\}, \ldots, \{i_{2n-1}, i_{2n}\}\}$. Order in each pair does not matter. For the moment we will allow matched pairs. Thus, for any given color i we have three possibilities: i is paired with itself; i is paired with j twice, for some $j \neq i$; or i is paired once with j and once with k, where i, j, and k are distinct. We notice that every color appears exactly twice in M.

We will say that i and j are *directly linked by* M if $\{i, j\} \in M$, and we will say that i and j are *linked by* M if we can get from i to j in a finite number of direct links. The relation of being linked is clearly symmetric and transitive, and since there are only finitely many colors, it must also be reflexive. Thus we have an equivalence relation on the set of colors $\{1, 2, \ldots, n\}$. Now let $P(M)$ be the corresponding set of equivalence classes; that is, $P(M)$ is the collection of subsets of $\{1, 2, \ldots, n\}$ where i and j are in the same subset if and only if they are linked by M. Thus, $P(M)$ is a partition of the set $\{1, 2, \ldots, n\}$.

As we saw in the example above, M induces a cyclic structure on the subsets in $P(M)$, by placing the colors that are directly linked by M adjacent to one another. That is, every subset of $P(M)$ corresponds to an element of S_n, the permutation group on n elements, written as a single cycle. For example, if $\{i, i\} \in M$, then $\{i\} \in P(M)$, corresponding to the 1-cycle (i). Likewise, if $\{i, j\}$ appears in M twice, then $\{i, j\} \in P(M)$, corresponding to the 2-cycle (ij). Finally, if $\{i_1, i_2\}, \{i_2, i_3\}, \ldots, \{i_k, i_1\} \in M$, $k \geq 3$, then $\{i_1, i_2, \ldots, i_k\} \in P(M)$, corresponding to *two distinct* k-cycles: $(i_1 i_2 \cdots i_k)$ and its inverse $(i_1 i_k i_{k-1} \cdots i_2)$. Thus, every set of sock pairs M corresponds to an element of S_n, written as a product of disjoint cycles. Likewise, since every permutation of S_n can be written as a product of disjoint cycles, every permutation in S_n corresponds to a set of sock pairs.

Let $C(a_1, a_2, \ldots, a_n)$ denote the number of permutations in S_n, written as a product of disjoint cycles, with a_i i-cycles for all $1 \leq i \leq n$. Notice that all of the a_i are nonnegative, and it has to be the case that $1a_1 + 2a_2 + \cdots + na_n = n$. To calculate

$C(a_1, a_2, \ldots, a_n)$ we first partition the set $\{1, 2, \ldots, n\}$ into a_1 subsets of size 1, a_2 subsets of size 2, and so on. The number of ways in which we can do this is

$$\binom{n}{1}\cdots\binom{n-a_1+1}{1}\frac{1}{a_1!}\binom{n-a_1}{2}\cdots\binom{n-a_1-2a_2+2}{2}\frac{1}{a_2!}\cdots$$
$$= \frac{n!}{a_1!\cdots a_n!\,(1!)^{a_1}\cdots(n!)^{a_n}}.$$

The factors $1/(a_i!)$ arise because, when we write a permutation as a product of disjoint cycles, the order in which the cycles appear does not matter. To complete the calculation of $C(a_1, a_2, \ldots, a_n)$ we need to make a cycle out of every subset in the partition. Recalling that there are $(j-1)!$ ways to make a cycle of length j from j distinct elements, we get

$$C(a_1, a_2, \ldots, a_n) = \frac{n!}{a_1!\cdots a_n!\,(1!)^{a_1}\cdots(n!)^{a_n}}(0!)^{a_1}\cdots((n-1)!)^{a_n}$$
$$= \frac{n!}{a_1!\cdots a_n!\,1^{a_1}\cdots n^{a_n}}.$$

Now let $C^*(a_1, a_2, \ldots, a_n)$ denote the number of ways to form a set of n sock pairs, corresponding to a permutation with a_i i-cycles. We saw above that for cycles of length 3 or greater, there are exactly two cycles that correspond to the same set of sock pairs. Therefore,

$$C^*(a_1, a_2, \ldots, a_n) = C(a_1, a_2, \ldots, a_n)\cdot\frac{1}{2^{a_3+\cdots+a_n}}.$$

After we have formed our sock pairs, distributing them is relatively easy. For the most part, one sock pair will be distinct from another. However, every time we have a cycle of length 2, we get two identical sock pairs. Thus, the number of ways we can distribute a set M of n sock pairs corresponding to a permutation with a_i i-cycles, $1 \le i \le n$, is $n!/2^{a_2}$.

To calculate u_n, we multiply $n!/2^{a_2}$ by $C^*(a_1, a_2, \ldots, a_n)$ and then sum over all possible nonnegative values $a_1, a_2, \ldots, a_n$ such that $1a_1 + 2a_2 + \cdots na_n = n$. To calculate d_n we do the same, but additionally we require that $a_1 = 0$, so that no sock is paired with one of the same color. Thus,

$$u_n = \sum \frac{(n!)^2}{a_1!\cdots a_n!\,1^{a_1}\cdots n^{a_n}2^{a_2+\cdots+a_n}} \tag{4}$$

and

$$d_n = \sum \frac{(n!)^2}{a_2!\cdots a_n!\,2^{a_2}\cdots n^{a_n}2^{a_2+\cdots+a_n}}. \tag{5}$$

In these sums, there is one term for each sequence $(a_1, \ldots, a_n)$ satisfying $a_1 + 2a_2 + 3a_3 + \cdots + na_n = n$.

The difficulty with this approach is in finding all possible nonnegative values $a_1, a_2, \ldots, a_n$ such that $1a_1 + 2a_2 + \cdots + na_n = n$. Below we work out the case when $n = 6$. For u_6 we sum all of the entries in the fourth column of the table, and for d_6 we exclude the rows where $a_1 \neq 0$.

This yields $u_6 = 202{,}410$ and $d_6 = 67{,}950$.

To conclude this section, we outline a recursive algorithm for finding all the n-tuples needed for the calculations of d_n and u_n. Actually, we show how to solve a

TABLE 1: Calculating u_6 and d_6

$(a_1, a_2, a_3, a_4, a_5, a_6)$	$C^*(a_1, a_2, a_3, a_4, a_5, a_6)$	$\frac{n!}{2^{a_2}}$	$C^*(a_1, a_2, a_3, a_4, a_5, a_6)\frac{n!}{2^{a_2}}$
(0, 0, 0, 0, 0, 1)	60	720	43200
(1, 0, 0, 0, 1, 0)	72	720	51840
(0, 1, 0, 1, 0, 0)	45	360	16200
(2, 0, 0, 1, 0, 0)	45	720	32400
(0, 0, 2, 0, 0, 0)	10	720	7200
(1, 1, 1, 0, 0, 0)	60	360	21600
(3, 0, 1, 0, 0, 0)	20	720	14400
(0, 3, 0, 0, 0, 0)	15	90	1350
(2, 2, 0, 0, 0, 0)	45	180	8100
(4, 1, 0, 0, 0, 0)	15	360	5400
(6, 0, 0, 0, 0, 0)	1	720	720

slightly more general problem. Let M and n be nonnegative integers. We wish to find all solutions $(a_1, a_2, \ldots, a_n)$ of nonnegative integers to the equation

$$1a_1 + 2a_2 + \cdots + na_n = M. \tag{6}$$

First, note that $a_i = 0$ whenever $i > M$. To organize the process, we break down the solutions in terms of the position of last nonzero entry.

Let $1 \le i \le n$ such that $a_i \neq 0$, and $a_j = 0$ for all $j > i$. (We will start with $i = n$ and work our way down to $i = 1$.) It follows that $1 \le a_i \le \lfloor \frac{M}{i} \rfloor$. More importantly, a_i can have any value in this range with a single exception: if $i = 1$, then $a_1 = M$ is the only solution. For $i > 1$, we fix a value for a_i, and let $m = M - ia_i$. (Note that $0 \le m < M$.) If we have $m = 0$, then all the remaining variables in the solution $(a_1, a_2, \ldots, a_n)$ must be zero. If $m > 0$, then we complete the solution by solving the equation

$$1a_1 + 2a_2 + \cdots + (i-1)a_{i-1} = m. \tag{7}$$

This, of course, is the recursive part. Since the number of variables in equation (7) is strictly smaller than the number of variables in equation (6), we know the process will end in a finite number of steps. After we have found all the solutions for each allowable value of a_i, we decrease i by one and start all over again.

Recurrence relations

Although we now have a solution, the difficulty of the calculations involved motivates us to try another approach. A common combinatorial strategy is to model problems with recurrence relations, and this tactic was used successfully by the Indian mathematicians Anand, Dumir, and Gupta in 1966 (see [**1**]). In this section, we present a streamlined version of their solution. Beginning with u_n, we divide into cases according to what happens to the last pair of socks.

Case 1. One of n people gets $\{n, n\}$; this can happen in nu_{n-1} ways.

Case 2. If pair n is split, there are two subcases to consider.

Subcase (i). We have two sock pairs of the form $\{i, n\}$ for some $1 \le i < n$; we choose 2 people to distribute them to, then distribute the remaining $n - 2$ pairs of socks. There are $\binom{n-1}{1}\binom{n}{2}u_{n-2}$ such distributions.

Subcase (ii). Pair n gets split into $\{i, n\}$ and $\{j, n\}$ for some $1 \le i < j < n$ and these get distributed to 2 people. Temporarily consider the remaining sock i and sock j as a matching pair. If we distribute these $n-2$ 'pairs' in u_{n-2} ways, we are undercounting by a factor of 2 those distributions in which i and j go to different people. By the reasoning in Case 1, i and j go to the same person in exactly $(n-2)u_{n-3}$ ways. Hence, there are $\binom{n-1}{2} P(n, 2)[2u_{n-2} - (n-2)u_{n-3}]$ such distributions.

After some algebraic simplification, we get the recurrence relation

$$u_n = nu_{n-1} + \frac{n(n-1)^2}{2}\left[(2n-3)u_{n-2} - (n-2)^2 u_{n-3}\right], \quad \text{for } n \ge 3.$$

We require three initial conditions; on the basis that "there is always one way to do nothing" (as one of our students puts it), we have $u_0 = 1$, $u_1 = 1$, $u_2 = 3$.

A similar analysis gives a recurrence relation for d_n. We can ignore Case 1, and in subcase (i) of Case 2, we simply replace u_{n-2} with d_{n-2}. The latter part of subcase (ii) requires some modification because d_{n-2} will not count distributions in which somebody gets the 'pair' $\{i, j\}$. We can count these separately by giving $\{i, j\}$ to one of $n-2$ people and then distributing the remaining pairs in d_{n-3} ways. The distributions in which $\{i, j\}$ is split number exactly $2d_{n-2}$. A bit of algebra yields

$$d_n = \frac{n(n-1)^2}{2}\left[(2n-3)d_{n-2} + (n-2)^2 d_{n-3}\right], \quad \text{for } n \ge 3.$$

The corresponding initial conditions are $d_0 = 1$, $d_1 = 0$, $d_2 = 1$.

Armed with these relations and *Maple*, we can easily generate values for u_n and d_n; we display the first ten in the table below.

TABLE 2: Some values of u_n and d_n

n	u_n	d_n
3	21	6
4	282	90
5	6210	2040
6	202,410	67,950
7	9,135,630	3,110,940
8	545,007,960	187,530,840
9	41,514,583,320	14,398,171,200
10	3,930,730,108,200	1,371,785,398,200

There are several things to notice about these numbers. First, the values for $n = 6$ agree with those in the previous section, happily. Second, like many combinatorial quantities, they grow amazingly quickly. Third, u_n is roughly three times d_n. More precisely, as n increases, further computations show that d_n/u_n approaches $1/e$, the same ratio that $D_n/n!$ approaches, although the former at a slower rate than the latter. This is the first of several intriguing connections to the derangement numbers.

The Online Encyclopedia of Integer Sequences contains both of these sequences, described somewhat differently: (u_n) is sequence $A000681$, which counts the number of $n \times n$ nonnegative integer matrices such that every row sums to 2 and every column sums to 2, and (d_n) is sequence $A001499$, which counts the number of $n \times n$, $0-1$ matrices with exactly two 1's in each row and each column. The connection becomes obvious if we formulate the sock problem graph theoretically. Define a bipartite graph $G = (X, Y, E)$ with $|X| = |Y| = n$; each vertex in X represents a matching sock pair,

and each vertex in Y represents a person. If each vertex has degree two, then everybody gets exactly two mismatched socks, and the incidence matrix is precisely an $n \times n$, $0-1$ matrix with exactly two 1's in each row and each column. The number of such incidence matrices is the solution to the Deranged Sock Problem. For the more general Sock Distribution Problem, we keep the requirement that G be 2-regular, but allow multiple edges.

Generating functions

Once we have recurrence relations for u_n and d_n, a natural impulse is to attempt to solve them to obtain non-recursive formulas, with the hope that they are simpler than the ones we found in our first approach. That the recurrence relations are nonlinear suggests that we try a generating function approach: That both u_n and d_n involve arrangements suggests that we use exponential generating functions. To this end, we define the formal power series

$$F(x) = \sum_{n=0}^{\infty} u_n \frac{x^n}{n!} \qquad \text{and} \qquad G(x) = \sum_{n=0}^{\infty} d_n \frac{x^n}{n!},$$

or equivalently,

$$F(x) = \sum_{n=0}^{\infty} \frac{u_n}{n!} x^n \qquad \text{and} \qquad G(x) = \sum_{n=0}^{\infty} \frac{d_n}{n!} x^n.$$

The point of this slight rewriting is to take advantage of a slight rewriting of equation (3) from the introduction. Since $P(n,k) = n!/(n-k)!$, we get

$$\frac{u_n}{n!} = \sum_{k=0}^{n} \binom{n}{k} \frac{d_{n-k}}{(n-k)!}.$$

If we let $\upsilon_n = u_n/n!$ and $\delta_n = d_n/n!$, then the equation above can be expressed as

$$\upsilon_n = \sum_{k=0}^{n} \binom{n}{k} \delta_{n-k}. \tag{8}$$

At this point, a notational trick from combinatorial analysis pays great dividends. When working with generating functions of the sequence $a_0, a_1, a_2, \ldots$, we can stipulate that $a^n \equiv a_n$. Then the ordinary and exponential generating functions of the sequence become, respectively,

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a^n x^n = \frac{1}{1-ax} \qquad \text{and} \qquad \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{a^n x^n}{n!} = e^{ax}.$$

Remarkably, all formal operations with power series carry through with the constant indices treated as powers. If we set $\upsilon^n \equiv \upsilon_n$ and $\delta^n \equiv \delta_n$, then equation (8) becomes

$$\upsilon^n = \sum_{nk=0}^{n} \binom{n}{k} \delta^n = (1+\delta)^n. \tag{9}$$

This yields $\sum \upsilon^n x^n = \sum (1+\delta)^n x^n$, which can be rearranged to produce

$$\frac{1}{1-\upsilon x} = \frac{1}{1-(1+\delta)x} = \frac{1}{(1-\delta x)-x} = \frac{1}{1-\delta x}\left[1 + \frac{x}{1-\upsilon x}\right].$$

By definition, the ordinary generating function of the υ_n is $F(x)$, the exponential generating function of the u_n; similarly, the ordinary generating function of the δ_n is $G(x)$. Thus we have shown

$$F(x) = G(x)\,[1 + xF(x)]\,.$$

As pretty as this relationship is, it is not clear how it can help us find non-recursive formulas for u_n and d_n. However, another way we can use (9) is to obtain

$$\sum_{n=0}^{\infty} \frac{\upsilon^n x^n}{n!} = \sum_{n=0}^{\infty} \frac{(1+\delta)^n x^n}{n!},$$

which simplifies to

$$e^{\upsilon x} = e^{(1+\delta)x} = e^x e^{\delta x}. \tag{10}$$

This gives a relationship between what we term the *double-exponential* generating functions,

$$f(x) = e^{\upsilon x} = \sum_{n=0}^{\infty} \frac{u_n}{n!}\frac{x^n}{n!} \qquad \text{and} \qquad g(x) = e^{\delta x} = \sum_{n=0}^{\infty} \frac{d_n}{n!}\frac{x^n}{n!}.$$

We can now observe another tantalizing connection to the derangement numbers. The exponential generating function of the D_n is known to be $D(x) = e^{-x}/(1-x)$. Since $1/(1-x)$ can be interpreted as the exponential generating function $U(x)$ of the permutation numbers, $n!$, we have $U(x) = e^x D(x)$, the same relationship as (10). This suggests that we investigate these double-exponential generating functions further.

Double-exponential generating functions

We now exploit our earlier analysis in terms of partitions and cyclic permutations. Recall that the number of permutations of n objects consisting of a_i i-cycles is given by

$$C(a_1, a_2, \ldots, a_n) = \frac{n!}{a_1! \cdots a_n!\, 1^{a_1} \cdots n^{a_n}}.$$

In [**4**], Riordan shows that a multivariable ordinary generating function for these numbers is

$$C_n(t_1, t_2, \ldots, t_n) = \sum \frac{n!}{a_1! \cdots a_n!} \left(\frac{t_1}{1}\right)^{a_1} \cdots \left(\frac{t_n}{n}\right)^{a_n},$$

where the sum is over all nonnegative integers $a_1, \ldots, a_n$ satisfying $a_1 + 2a_2 + \cdots + na_n = n$. Riordan further demonstrates that

$$\sum_{n=0}^{\infty} C_n(t_1, \ldots, t_n)\frac{x^n}{n!} = \exp\left(t_1\frac{x}{1} + t_2\frac{x^2}{2} + t_3\frac{x^3}{3} + t_4\frac{x^4}{4} + \cdots\right).$$

Expressing the results (4) from our earlier approach in this new notation, we obtain

$$\frac{u_n}{n!} = C_n\left(1, \frac{1}{2}, \ldots, \frac{1}{2}\right) \qquad \text{and} \qquad \frac{d_n}{n!} = C_n\left(0, \frac{1}{2}, \ldots, \frac{1}{2}\right),$$

which in turn leads to

$$\begin{aligned} f(x) &= \exp\left(x + \frac{1}{2}\left[\frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots\right]\right) \\ &= \exp\left(\frac{1}{2}\left[\log\frac{1}{1-x} + x\right]\right) = \frac{e^{x/2}}{\sqrt{1-x}}; \\ g(x) &= \exp\left(\frac{1}{2}\left[\frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots\right]\right) \\ &= \exp\left(\frac{1}{2}\left[\log\frac{1}{1-x} - x\right]\right) = \frac{e^{-x/2}}{\sqrt{1-x}}. \end{aligned}$$

(An alternate derivation of these formulas appears in section 8 of [**1**].)

We can now use the closed expressions of the double-exponential generating functions to derive another set of non-recursive formulas for u_n and d_n. Recall the binomial series expansion

$$\frac{1}{\sqrt{1-x}} = (1-x)^{-1/2} = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-x)^n,$$

where $\binom{-1/2}{0} = 1$ and for $n \geq 1$,

$$\begin{aligned} \binom{-1/2}{n} &= \frac{(-1/2)(-3/2)(-5/2)\cdots(-1/2-n+1)}{n!} \\ &= \frac{(-1)^n 1\cdot 3\cdot 5\cdots(2n-1)}{2^n n!}. \end{aligned}$$

Observe that this series can be viewed as the exponential generating function of the sequence $a_n = [1\cdot 3\cdot 5\cdots(2n-1)]/2^n$. Using the formal procedure for multiplying exponential series, and remembering that $f(x)$ and $g(x)$ are *double*-exponential generating functions, we obtain

$$\begin{aligned} u_n &= \frac{n!}{2^n}\left(1 + \sum_{k=1}^{n} \binom{n}{k} 1\cdot 3\cdot 5\cdots(2k-1)\right), \\ d_n &= \frac{n!}{2^n}\left((-1)^n + \sum_{k=1}^{n} \binom{n}{k} 1\cdot 3\cdot 5\cdots(2k-1)(-1)^{n-k}\right). \end{aligned}$$

Asymptotic behavior

We now have both recursive and non-recursive formulas for u_n and d_n, but neither shed much light on the observation that, just like $D_n/n!$, the ratio d_n/u_n approaches $1/e$ as n increases. To gain insight into this behavior, we must delve more deeply into the theory of generating functions.

By merely replacing the variable x with z, we can mentally transform a formal power series into a function of complex variables, which must be analytic on some disk in the complex plane centered at 0. In particular, $f(z) = e^{z/2}/\sqrt{1-z}$ and $g(z) = e^{-z/2}/\sqrt{1-z}$ are each analytic in the unit disk, with a single algebraic singularity at $z_0 = 1$. Fortunately, some heavy machinery from complex analysis exists for exactly this sort of function. The following can be found in [**6**].

THEOREM 1. (DARBOUX'S LEMMA) *Let $v(z)$ be analytic in some disk $|z| < 1+\eta$ where $\eta > 1$. Let $\beta \in \mathbb{R}\backslash\{0, 1, 2, \ldots\}$. Suppose that in a neighborhood of $z = 1$, $v(z) = \sum v_n(1-z)^n$. Then for every integer $m \geq 0$, the coefficient of z^n in the expansion of $v(z)(1-z)^\beta$ is*

$$\left[\sum_{k=0}^{m} \binom{\beta+k}{n}(-1)^n v_k\right] + O(n^{-m-\beta-2}).$$

To apply this to $f(z)$, we use $v(z) = e^{z/2}$. This is an entire function; its Taylor expansion centered at $z = 1$ is

$$e^{z/2} = \sum_{n=0}^{\infty} \frac{v^{(n)}(1)}{n!}(z-1)^n = \sum_{n=0}^{\infty} \frac{(-1)^n e^{1/2}}{2^n n!}(1-z)^n.$$

Then using Darboux's Lemma with $m = 2$, we obtain

$$\frac{u_n}{(n!)^2} = e^{1/2}\left[\binom{-1/2}{n} + \frac{1}{2}\binom{-3/2}{n} + \frac{1}{8}\binom{-5/2}{n}\right] + O(n^{-7/2}). \tag{11}$$

For large values of n, we can simplify the fractional binomial coefficients in this expression with this result, also in [**6**].

THEOREM 2. *Let $\alpha \in \mathbb{R}\backslash\{0, 1, 2, \ldots\}$. Then as $n \to \infty$,*

$$\binom{\alpha}{n} \sim \frac{(-1)^n n^{-\alpha-1}}{\Gamma(-\alpha)}.$$

The appearance of the gamma function need not overly alarm us, as we can compute all the values we need from the simple pair of properties $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(z)\Gamma(1-z) = \pi/\sin \pi z$ (see [**3**]). Substituting into (11), we get for large values of n,

$$u_n \sim \frac{(n!)^2 e^{1/2}}{\sqrt{n\pi}}\left[1 + \frac{1}{4n} + \frac{3}{32n^2}\right].$$

Replacing $v(z) = e^{z/2}$ with $w(z) = e^{-z/2}$, we get an asymptotic formula for d_n:

$$d_n \sim \frac{(n!)^2 e^{-1/2}}{\sqrt{n\pi}}\left[1 - \frac{1}{4n} + \frac{3}{32n^2}\right].$$

The table below illustrates how accurate these approximations are, even for small values of n.

TABLE 3: Asymptotic values of u_n and d_n.

n	u_n	$\frac{(n!)^2 e^{1/2}}{\sqrt{n\pi}}\left[1+\frac{1}{4n}+\frac{3}{32n^2}\right]$	d_n	$\frac{(n!)^2 e^{-1/2}}{\sqrt{n\pi}}\left[1-\frac{1}{4n}+\frac{3}{32n^2}\right]$
5	6210	6312.3	2040	2101.8
10	3.931×10^{12}	3.974×10^{12}	1.371×10^{12}	1.391×10^{12}
20	1.239×10^{36}	1.247×10^{36}	4.444×10^{35}	4.474×10^{35}
30	1.200×10^{64}	1.205×10^{64}	4.341×10^{63}	4.360×10^{63}
40	9.823×10^{94}	9.853×10^{94}	3.568×10^{94}	3.580×10^{94}
50	1.220×10^{128}	1.223×10^{128}	4.443×10^{127}	4.454×10^{127}

The asymptotic formulas finally clear up the mystery:

$$\lim_{n\to\infty}\frac{d_n}{u_n}=\lim_{n\to\infty}\frac{\frac{(n!)^2e^{-1/2}}{\sqrt{n\pi}}\left[1-\frac{1}{4n}+\frac{3}{32n^2}\right]}{\frac{(n!)^2e^{1/2}}{\sqrt{n\pi}}\left[1+\frac{1}{4n}+\frac{3}{32n^2}\right]}=\frac{1}{e}.$$

Acknowledgment We would like to thank the students of Hamilton's 2005 combinatorics class for their useful discussions on this problem.

REFERENCES

1. Harsh Anand, Vishwa Chander Dumir, and Hansraj Gupta, A combinatorial distribution problem, *Duke Mathematical Journal* **33** (1966) 757–769. http://dx.doi.org/10.1215/S0012-7094-66-03391-6
2. Kenji Mano, On the formula of ${}_nH_r$, *Scientific Reports of the Faculty of Literature and Science: Hirosaki University* **8** (1961) 58–60.
3. J. E. Marsden and M. J. Hoffman, *Basic Complex Analysis*, W. H. Freeman, New York, 1999.
4. John Riordan, *Introduction to Combinatorial Analysis*, Dover, Mineola, NY, 2002.
5. Alan Tucker, *Applied Combinatorics*, John Wiley, New York, 2007.
6. Herbert S. Wilf, *Generatingfunctionology*, A. K. Peters, Natick, MA, 2006.

Summary It is an elementary combinatorial problem to determine the number of ways n people can each choose two gloves from a pile of n distinct pairs of gloves, with nobody getting a matching pair. Change the gloves to socks (with right socks being indistinguishable from left socks), however, and the problem becomes surprisingly more difficult. We show how this problem can be solved using a wide range of discrete mathematics tools: the principle of inclusion-exclusion; partitions; cyclic permutations; recurrence relations; as well as both ordinary and exponential generating functions. We even draw on a result from complex analysis to show that the fraction of all sock distributions that are deranged in this sense converges to $1/e$.

SALLY COCKBURN is an associate professor in the mathematics department at Hamilton College, with interests in combinatorics, graph theory, linear optimization and set theory. For relief from mathematical dilettantism, she plays squash and tries to keep up with her menagerie of dogs.

JOSHUA LESPERANCE is currently a visiting assistant professor at Oberlin College, and his mathematical interests are varied and ever-evolving. He received a B.S. from Rochester Institute of Technology and both an M.S. and Ph.D. from the University of Notre Dame. In his free time, he enjoys playing the guitar and hiking with his wife and two Siberian Huskies.

About the Cover

From deranged socks to deranged sock monkeys is no big leap. If you are familiar with those iconic stuffed animals made from knit cotton work socks, buttons, yarn, and snips of cloth, you'll know that no two are alike and each has a very specific personality. Some look happy, some look sad, some look devilish and, yes, some look deranged.

Collector Ron Warren certainly knows about sock monkeys. He has over 1800 of them. He and photographer Arne Swenson have documented his collection in a book called, of course, *Sock Monkeys,* published by Ideal World Books. Artist Annie Stromquist's linocut and watercolor portrait on this month's cover was inspired by the book.

From Bocce to Positivity: Some Probabilistic Linear Algebra

KENT E. MORRISON
American Institute of Mathematics, Palo Alto, CA 94306
California Polytechnic State University, San Luis Obispo, CA 93407
kmorriso@calpoly.edu

The questions

At the American Institute of Mathematics, we often play a lunchtime game of bocce in a nearby park. Each round begins with the winner of the previous round tossing the pallino, a small white marker ball, out onto the lawn somewhere. Then each player takes turns throwing larger balls as close as possible to the pallino. Winners like to think their good fortune is a result of skill, but we see a lot of variability in the results leading us to conclude that chance plays a major role. Usually the balls cluster around the pallino, but from time to time they do not, and once, when all eight of them were off to one side—so that the pallino was not contained in their convex hull—I became intrigued with the possibility of figuring out the probability of that occurring.

Answering this question leads naturally to other questions that can be answered with the same approach. One is about positive solutions of systems of linear equations. We might be especially interested in positive solutions because the equations are a model for some real-world problem where negative values of the variables do not make sense. Suppose that a system of liner equations is drawn out of a hat—that is, the coefficients of the system are random. What is the probability that there is a positive solution?

A third question concerns random two-person zero-sum games. The payoff matrix for such a game is an $m \times n$ rectangular matrix of real numbers. One player chooses a row, the other player a column, and the column player pays the row player the amount of the corresponding matrix entry—a negative number meaning that the row player actually pays the column player. Most of the time, players do not have unique best choices, but they do have optimal probabilistic strategies specifying the probabilities for choosing among their options. It stands to reason that the player having more options is more likely to have the advantage, but can we quantify that? More precisely, assuming the game matrix has random entries as likely to be positive as negative, what is the probability that the row player expects a positive payoff?

Some low-dimensional situations

For each question some special cases are easy to answer, and we begin with those.

(1) Bocce, as it is played in this world, is a two-dimensional game, but imagine playing bocce in d dimensions for any $d \geq 1$. The special case of $d = 1$ we'll take care of now. Place the pallino at the origin on the real line, and assume that the n players' balls are independent random points with equal probability to be on either side of the origin and zero probability to be exactly at the origin. Then the origin fails to be in the convex hull of the random points exactly when all the points are on the

Math. Mag. **86** (2013) 110–119. doi:10.4169/math.mag.86.2.110.

same side of the origin. The first point can be on either side with the remaining $n-1$ points on the same side as the first point, an event that has probability $1/2^{n-1}$.

(2) Let's start with a single linear equation in two variables,

$$a_1x_1 + a_2x_2 = b,$$

so that the set of solutions is a line in the plane—as long as a_1 and a_2 are not both 0. Assume that a_1 and a_2 are independent random variables, each having probability $1/2$ of being positive, probability $1/2$ of being negative, and probability zero of being 0. When $b \neq 0$, there are two intercepts, $x_1 = b/a_1$ and $x_2 = b/a_2$, and four possibilities for the signs of the intercepts, each having probability $1/4$. Unless both intercepts are negative, the line will meet the first quadrant and there will be positive solutions. Therefore, the probability of a positive solution is $3/4$. However, when $b = 0$, the solution set is a line through the origin with slope $-a_1/a_2$. The sign of the slope is equally likely to be positive as negative, and so the probability is $1/2$ that the line meets the first quadrant, giving a positive solution.

A similar analysis works for one equation in n variables,

$$a_1x_1 + \cdots + a_nx_n = b.$$

For $b \neq 0$, the probability of a positive solution is $1 - 1/2^n$, while for $b = 0$ the probability is $1 - 1/2^{n-1}$.

(3) Let's consider the game with $m = 1$ and arbitrary n. The row player has only one choice, and the column player just picks the smallest entry since he wants to minimize what he pays. The minimum value will be positive when all n entries are positive. Hence, if we assume that the entries are independent and just as likely to be positive as negative and with no chance of being zero, then the probability that they are all positive is $1/2^n$, and that is the probability that the row player has the advantage.

Two-dimensional bocce

We return to the original bocce game as it is played on a two-dimensional surface. The pallino is at the origin and the players' balls are random points $z_1, \ldots, z_n$ in $\mathbb{R}^2$. In actual play, especially with skilled players, the locations of the balls are not independent, because players attempt to knock opponents' balls out of the way or to set up blocking positions in front of the pallino. However, our casual game is not played on a smooth bocce court but rather in a park with bumpy terrain, patches of dead lawn, trees, hills, sidewalks, and playground equipment. The player with the nearest ball on each round begins the next round by throwing the pallino wherever he or she chooses, and so all the irregularities of the terrain make it reasonable to assume that the location of the players' balls are independent random points. The second key assumption is that the probability distribution for each ball has a density function symmetric with respect to the origin. For a probability density function f, this means that $f(-z) = f(z)$ for $z \in \mathbb{R}^2$. (We are not requiring that the random points be identically distributed.) Common examples include bivariate normal distributions centered at the origin and uniform distributions of regions such as disks or rectangles centered at the origin. It follows from this assumption that for any line through the origin, the probability is $1/2$ to be in each of the open half-planes on either side of the line.

With our assumptions in place, let's define E to be the event that the origin does not lie within the convex hull of n random points $z_1, \ldots, z_n$. For such a configuration, there is a unique distinguished point among the n points with the property that all the

remaining points are in the half-plane described by starting at that point and going π radians counterclockwise. This shows that E is the disjoint union of events $E_1, \ldots, E_n$, where E_i is the event that z_i is the distinguished point. Therefore,

$$\mathrm{P}(E) = \sum_{i=1}^{n} \mathrm{P}(E_i),$$

but $\mathrm{P}(E_i) = 1/2^{n-1}$ because z_i can be anywhere and the other $n-1$ points must be in the correct half-plane. Therefore,

$$\mathrm{P}(E) = \frac{n}{2^{n-1}}.$$

This result is consistent with the informal observations in our bocce games in the park. We usually have four players, each throwing two balls, and so $n = 8$, in which case $\mathrm{P}(E) = 8/2^7 = 1/16$.

Higher-dimensional bocce

The two-dimensional analysis does not seem to work in three dimensions because there is no natural way to associate to each nonzero point in $\mathbb{R}^3$ a half-space with that point on the separating plane. The same goes for higher dimensions. But in 1962, J. G. Wendel [**7**] found an elegant solution relying fundamentally on a much older theorem of L. Schläfli, which counts the number of regions in $\mathbb{R}^d$ created by n generic hyperplanes through the origin. Schläfli's result is in *Theorie der vielfachen Kontinuität*, written between 1850 and 1852, which is one of the seminal contributions to the development of higher-dimensional geometry in the nineteenth century. Despite repeated efforts by Schläfli and others, it was almost 50 years before it was eventually published in 1901, six years after his death. Schläfli's result can be found in the free Google Books edition [**5**, p. 41] or in his collected works [**6**, p. 211].

As Wendel puts it, there are n points "scattered at random on the surface of the unit sphere" in $\mathbb{R}^d$, and the problem is to evaluate the probability that all the points lie on some hemisphere. But this is just what we want, because the origin is not in the convex hull of nonzero points $z_1, \ldots, z_n$ in $\mathbb{R}^d$ if and only if the points all lie in some half-space, or equivalently that $z_1/|z_1|, \ldots, z_n/|z_n|$ all lie on some hemisphere of the unit sphere.

What follows is a modified version of Wendel's solution. Let $p(n, d)$ be the probability that the convex hull of n random points in $\mathbb{R}^d$ does not contain the origin (equivalently, that the points lie in some half-space). We assume that the points $z_1, \ldots, z_n$ are independent and the probability distributions of the points are symmetric with respect to the origin and that they have density functions. For each of the 2^n sign vectors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$, we define a random variable X_ε. The value of X_ε is 1 if the points $\varepsilon_1 z_1, \ldots, \varepsilon_n z_n$ all lie in a half-space; otherwise the value is 0. Then $p(n, d) = E(X_{\mathbf{1}})$, where $\mathbf{1} = (1, \ldots, 1)$. Because the distributions are symmetric with respect to the origin, $E(X_\varepsilon)$ is independent of ε, and so $2^n E(X_{\mathbf{1}}) = \sum_\varepsilon E(X_\varepsilon) = E\big(\sum_\varepsilon X_\varepsilon\big)$. Therefore,

$$p(n, d) = \frac{1}{2^n} E\left(\sum_\varepsilon X_\varepsilon\right).$$

The next step is showing that the sum $\sum_\varepsilon X_\varepsilon$ is constant almost surely, and that the value of the constant is the number of connected regions in $\mathbb{R}^d$ created by the

n hyperplanes through the origin that are orthogonal to the z_i. Each of the regions corresponds to a sign vector ε, which describes that region as a particular intersection of half-spaces. That is, two points v and w are in the same region if and only if the inner products $\langle v, z_i\rangle$ and $\langle w, z_i\rangle$ have the same sign for $i = 1, \ldots, n$. But not all sign vectors correspond to regions because the intersection of half-spaces described by a sign vector can be empty. Now for each ε that does come from a region, let v be a point in the region. Then $\varepsilon_1 z_1, \ldots, \varepsilon_n z_n$ all lie in the half-space of points x such that $\langle v, x\rangle > 0$, and so $X_\varepsilon = 1$ for the random points $z_1, \ldots, z_n$. Summing over ε, we conclude that $\sum_\varepsilon X_\varepsilon$ is equal to the number of regions. (There are configurations of the points for which the sum $\sum_\varepsilon X_\varepsilon$ does not achieve this value but is something less. This occurs when there is some unexpected linear dependence among the z_i. To be precise, the exceptional configurations for $n \le d$ are those for which $z_1, \ldots, z_n$ are linearly dependent; for $n > d$, they are those for which some d of the points are linearly dependent. For example, if z_1, z_2, z_3 in $\mathbb{R}^3$ all lie in a plane, then the planes normal to them divide space into only six regions rather than eight.)

The last ingredient we need is Schläfli's formula for the number of regions. Letting $r(n, d)$ denote the number of regions created by n hyperplanes through the origin in $\mathbb{R}^d$, what Schläfli proved is that

$$r(n, d) = 2\sum_{j=0}^{d-1} \binom{n-1}{j}.$$

How could you discover this formula? Well, $r(n, d)$ satisfies the recurrence relation

$$r(n, d) = r(n-1, d) + r(n-1, d-1)$$

with boundary conditions $r(n, 2) = 2n$ and $r(n, d) = 2^n$ for $n \le d$. The boundary conditions are straightforward, but the recurrence relation is more subtle and is explained in the appendix. With this information you can compute several values of $r(n, d)$ and then hope to notice that the difference $r(n, d) - r(n, d-1)$ is always twice a binomial coefficient; in fact, it is $2\binom{n-1}{d-1}$. From that you build the formula, which can then be proved rigorously by showing it satisfies the recurrence relation and boundary conditions.

With this we have everything we need to see that

$$p(n, d) = \frac{1}{2^n} r(n, d) = \frac{1}{2^{n-1}} \sum_{j=0}^{d-1} \binom{n-1}{j}.$$

TABLE 1: Values of $p(n, 3)$ for $1 \le n \le 13$

1	2	3	4	5	6	7	8	9	10	11	12	13
1.0	1.0	1.0	.875	.686	.500	.344	.227	.145	.090	.055	.033	.019

With a little more work we can see that the location of the first ball does not really matter for the result. That is, you can specify the point z_1 while the other points are random, and the probability that they all lie in a half-space is still $p(n, d)$. To understand this, consider the sum of the 2^{n-1} random variables X_ε, where ε ranges over the sign vectors in $\{\pm 1\}^n$ such that $\varepsilon_1 = 1$. Then this sum is equal to the number of sign vectors corresponding to the regions that have $\varepsilon_1 = 1$. But there is a one-to-one correspondence between those regions and the regions whose sign vectors have $\varepsilon_1 = -1$, where the correspondence just pairs a region R with its negative $-R$. Thus, the sum we

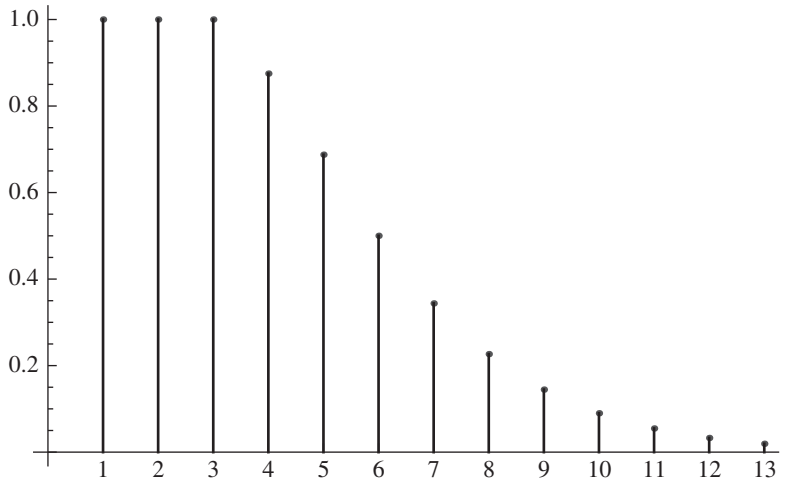


Figure 1 The probability $p(n, 3)$ for $1 \leq n \leq 13$

want is $(1/2)r(n, d)$. Now, we divide this by 2^{n-1} to get the expectation of $X_\mathbf{1}$, which is the probability that we want, but this gives us exactly $p(n, d) = (1/2^n)r(n, d)$.

We end this section with the intriguing observation that

$$p(n, d) + p(n, n - d) = 1.$$

That is, $p(n, d)$ and $p(n, n - d)$ are complementary probabilities. This identity holds for $n \geq 0$ and all integer values of d with the understanding that $p(n, d) = 0$ for $d \leq 0$. Although it is easy to verify algebraically, we will return to it with a geometric proof in the section on random subspaces.

Positive solutions of linear equations

We can use what we already know about convex hulls to find the probability that a random system of linear equations has a positive solution. The key is that the existence of a positive solution for a homogeneous system of linear equations is equivalent to the property that the origin is in the convex hull of the column vectors of the coefficient matrix. Here are the details.

Consider an $m \times n$ matrix A and the system of linear equations written as $Ax = 0$. If x is a solution and $z_1, \ldots, z_n \in \mathbb{R}^m$ are the columns of A, then $\sum x_i z_i = 0 \in \mathbb{R}^m$. Furthermore, if x is a *positive* vector (meaning that $x_i \geq 0$ for all i and at least one of the x_i is positive), then we can multiply x by the scalar $1/\sum x_i$ to get a solution vector t with the property that $t_i \geq 0$ and $\sum t_i = 1$, and thus 0 is in the convex hull of the z_i.

Therefore, if the columns of the matrix A are random points in $\mathbb{R}^m$ that satisfy the assumptions in the bocce problem, then the probability that $Ax = 0$ has a positive solution is the probability that n points in $\mathbb{R}^m$ contain the origin in their convex hull, which is $1 - p(n, m) = p(n, n - m)$. Those assumptions are satisfied if the entries of A are independent random variables distributed with probability densities that are even functions. Examples include normal distributions and uniform distributions in balanced intervals of the form $[-c, c]$.

Turning to the random system of equations $Ax = b$, where $b \neq 0$, we may consider b as a random vector or as a fixed vector, because both lead to the same probability that there is a positive solution. Now suppose that there is a positive solution x, and let

$z_1, \ldots, z_n$ be the columns of A. Then $\sum x_i z_i = b$. Move b to the other side and scale by $(1 + \sum x_i)^{-1}$ to see that 0 is a convex combination of $-b$ and the z_i. Now we are in the situation of $n + 1$ points with one fixed and the rest random, and in the previous section we determined that the probability that they lie in a half-space is $p(n + 1, m)$. Therefore, the complementary probability $1 - p(n + 1, m) = p(n + 1, n + 1 - m)$ is the probability that the origin is in the convex hull. When the origin is in the convex hull we have an expression of the form

$$-t_0 b + t_1 z_1 + \cdots + t_n z_n = 0,$$

with $t_i \geq 0$, and we can move b back to the right side to show a positive solution for $Ax = b$ as long as $t_0 \neq 0$.

Let's consider the possibility that $t_0 = 0$ in such a convex combination. Then 0 is in the convex hull of the z_i alone. Because any m of the z_i are linearly independent and $n > m$, the convex hull does not lie in a lower-dimensional subspace and so it has nonempty interior. Since the probability is zero that the origin is on the boundary of the convex hull, it must be in the interior, and so there is an open ball containing the origin and lying within the convex hull. For a sufficiently small $\lambda > 0$, the point λb is in the ball and hence in the convex hull of the z_i. Therefore, there exist $s_i \geq 0$ such that $s_1 z_1 + \cdots + s_n z_n = \lambda b$. Multiplying both sides by $1/\lambda$ gives a positive solution of $Ax = b$.

In summary, the probability of a positive solution of $Ax = 0$ is $p(n, n - m)$. The probability of a positive solution of $Ax = b$, where b is a fixed or random nonzero vector, is $p(n + 1, n + 1 - m)$. As a partial check, let's look back at the case of one equation and two unknowns ($m = 1$ and $n = 2$) for which $p(n, n - m) = p(2, 1) = 1/2$ and $p(n + 1, n + 1 - m) = p(3, 2) = 3/4$. Those values are just what we found earlier. In general, just as in this special case, it is more likely that there is a positive solution when $b \neq 0$.

Random subspaces

Since the solutions of the equation $Ax = 0$ form the null space of the matrix A, the question about positive solutions becomes a question about the probability that a random subspace of $\mathbb{R}^n$ contains a positive vector. We can assume that the subspaces have a fixed dimension $n - m$, because the assumptions about the distribution of the entries of A imply that $\text{rank}(A) = m$, almost surely, so that the null space has dimension $n - m$.

The orthogonal complement to the null space of A is the row space of A or the range of A^t, and so it is a subspace of dimension m. That is, a random m-dimensional subspace of $\mathbb{R}^n$, for $m < n$, is generated by taking the span of m random vectors in $\mathbb{R}^n$. With reasonable assumptions, the vectors will be linearly independent with probability one, thus giving a span of dimension m.

What is the probability that a random m-dimensional subspace of $\mathbb{R}^n$ contains a positive vector? We'll see that the answer is $p(n, m)$. Let the random subspace be the row space of the random $m \times n$ matrix A as before (the columns are independent random vectors having probability densities symmetric with respect to the origin). In particular, the individual entries in A could be independent random variables whose probability density functions are even.

Now we apply the theorem from linear algebra known as Gordan's Theorem of the Alternative [**4**]: A subspace V of $\mathbb{R}^n$ contains a positive vector if and only if its orthogonal complement $V^\perp$ does not contain a strictly positive vector. Let V be the row space of A so that $V^\perp$ is the null space of A. Then the probability that $V^\perp$ does

not contain a strictly positive vector is the same as the probability that it does not contain a positive vector, namely, $1 - p(n, n-m)$, which is $p(n, m)$.

This, then, gives us a geometric explanation for the complementary probability identity mentioned earlier,

$$p(n, m) + p(n, n-m) = 1.$$

We can understand the result as a *probabilistic theorem of the alternative*: With probability one, a random subspace either contains a positive vector or its orthogonal complement contains a positive vector, but not both. In short, complementary subspaces define complementary events.

FIGURE 2 shows that for large n, the value of $p(n, m)$ rises quickly from nearly 0 to nearly 1 as m passes through $n/2$. As n increases, the plot becomes more and more like a step function. Note that for n even, $p(n, n/2)$ is always $1/2$.

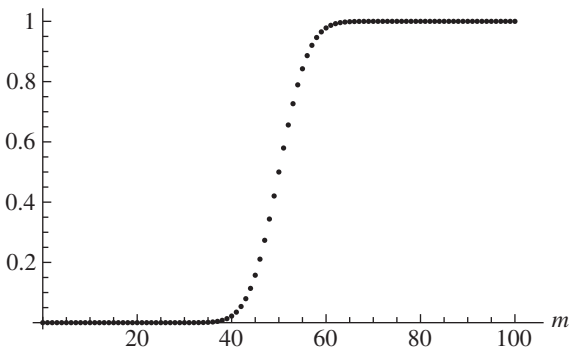


Figure 2 $p(n, m)$ for $n = 100$

We end this section with an explanation for the individual binomial probabilities

$$\frac{1}{2^{n-1}}\binom{n-1}{j}$$

that we see in the sums for $p(n, m)$. This, of course, is the probability of exactly j heads occurring in a sequence of $n-1$ tosses of a fair coin. Now instead of flipping a coin, we generate a sequence of n independent random vectors $v_1, \ldots, v_n$ in $\mathbb{R}^n$. Let V_m be the span of the first m vectors. Then V_m is a random subspace of dimension m. The probability that V_m contains a positive vector while V_{m-1} does not is

$$p(n, m) - p(n, m-1) = \frac{1}{2^{n-1}}\binom{n-1}{m-1}.$$

In other words, the integer-valued random variable whose value is the least m such that V_m contains a positive vector, has the same distribution as the number of heads in $n-1$ tosses of a fair coin.

Random games

Consider a random two-person zero-sum game described by an $m \times n$ matrix A. The rows represent the pure strategies of the row player and the columns the pure strategies of the column player, and the convention is that when the row player chooses row i and the column player chooses column j, the result is that the column player pays the row

player the amount a_{ij}. (If a_{ij} is negative, then the column player receives $-a_{ij}$ from the row player.) Thus, positive entries are good for the row player and negative entries are good for the column player. We assume that the entries are independent random variables with probability density functions that are even, so that each individual entry favors neither player.

In general, there is no optimal pure strategy for each player, but there are optimal mixed strategies. A mixed strategy is a probability distribution on the finite set of pure strategies. For the row player it is a probability vector $p = (p_1, \ldots, p_m)$, where p_i is the probability of choosing row i to play, and for the column player a mixed strategy is a probability vector $q = (q_1, \ldots, q_n)$, where q_j is the probability of playing column j. The players choose their strategies independently and so the expected payoff to the row player is

$$\sum_{i,j} a_{ij} p_i q_j,$$

which can be written as the product $p^t Aq$, where p and q are treated as column vectors.

The row player wants to choose p to make this product as large as possible, while the column player wants to choose q to minimize it. The Minimax Theorem, von Neumann's fundamental result, asserts that

$$\max_p \min_q p^t Aq = \min_q \max_p p^t Aq,$$

and this number is called the *value* of the game. Furthermore, the theorem asserts that there exist *optimal* mixed strategies p_* and q_*, not necessarily unique, such that ${p_*}^t Aq_*$ equals the value of the game. (For full treatment of this material, see the book by Guillermo Owen [**3**] or the e-book by Tom Ferguson [**2**].)

The game favors the row player when the value is positive, since the value of the game is the expected amount that the row player receives when the players use their optimal strategies. We'd like to know the probability of that event for a random m by n game. Intuition suggests that the game is more likely to favor the player with the greater number of strategies, and that for $m = n$ it should be equally likely that the game value is positive or negative.

In 1966 Thomas Cover [**1**] proved—with reasonable assumptions on the entries of the payoff matrix—that the probability that the game favors the row player is $p(m + n, m)$. To get an idea of how much the advantage is for the player with more strategies, consider the situation in which the column player has twice as many pure strategies as the row player, i.e., $n = 2m$. FIGURE 3 plots $p(n + m, m)$ vs. m, showing how quickly this probability approaches 0 as m increases.

To prove this result we assume that the entries of the payoff matrix A are independent random variables whose probability density functions are symmetric with respect to 0. For the value of the game to be positive, the row player must have a mixed strategy that gives a positive payoff no matter which column the opponent chooses. That is equivalent to the existence of a positive vector $x \in \mathbb{R}^m$ such that $\langle x, z_j \rangle > 0$ for $j = 1, \ldots, n$, where $z_j \in \mathbb{R}^m$ is the jth column of the payoff matrix A. (From such an x we get the mixed strategy by scaling it to make it a probability vector.)

If we multiply any of the rows and columns of the random A by -1, the probability that the resulting matrix favors the row player does not change because of the symmetry of the matrix entries. Changing the signs of rows is the same as multiplying A on the left by a diagonal matrix C with diagonal entries ± 1, while changing the signs of columns is equivalent to multiplying on the right by a diagonal matrix D with ± 1 entries.

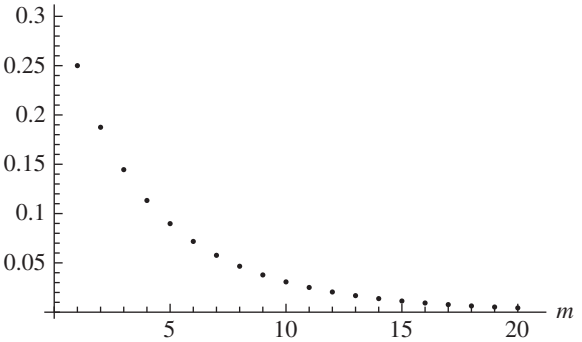


Figure 3 Probability that row player wins with m strategies when the column player has $2m$ strategies.

Let $W_{C,D}$ be the random variable equal to 1 if the game with payoff matrix CAD has positive value (i.e., favors the row player) and 0 otherwise. The probability we want is then the expectation $E(W_{I,I})$, and we have just noted that $E(W_{C,D})$ is independent of C and D.

Next, we show that the sum of $W_{C,D}$ is constant (with probability one) where C and D range over all pairs of diagonal ± 1 matrices. Consider the m coordinate hyperplanes (i.e., the planes orthogonal to the standard basis vectors) along with the n hyperplanes orthogonal to the z_j. These $m + n$ hyperplanes separate $\mathbb{R}^m$ into $r(m + n, m)$ regions by Schläfli's result. For x in one of the regions, let c_i be the sign (± 1) of x_i and let d_j be the sign of $\langle x, z_j \rangle$. Let C and D be the diagonal matrices with diagonal entries c_i and d_j. Then $x^t C$ is a strictly positive vector whose inner product with each column of CAD is positive, and thus the payoff matrix CAD describes a game that favors the row player. Therefore, with probability one, the random variable $\sum W_{C,D}$ is equal to the constant $r(m + n, m)$, and so $E(\sum W_{C,D}) = r(m + n, m)$. But $E(\sum W_{C,D}) = \sum E(W_{C,D}) = 2^{m+n} E(W_{I,I})$, and from this it follows that

$$E(W_{I,I}) = \frac{1}{2^{m+n}} r(m + n, m) = p(m + n, m).$$

Positive input, positive output

A linear system described by the $m \times n$ matrix A maps an input vector x in $\mathbb{R}^n$ to an output vector Ax in $\mathbb{R}^m$. Again the linear system may only make sense as a physical model when there exists some positive input x whose resulting output Ax is also positive. Now suppose that A is a random matrix with the same conditions on the independent entries as before. What is the probability that there is a positive input with positive output?

It turns out that we already have the answer to that question from the game theory situation, although in transposed form. Asking for a strictly positive output, rather than just a positive output, doesn't change the probability. For Ax to be strictly positive means that the inner product of x with each row of A is positive. If x is also positive, then we have the conditions for the game with matrix A^t to have a positive value, and so we conclude that the probability of that occurring is $p(n + m, n)$.

An easy case to check directly is $m = 1$. Then the linear system is given by the map $x \mapsto \sum a_i x_i$. As long as any $a_i > 0$, it is possible to make x_i a large positive number and the other x_j small positive numbers so that the sum is positive. The complementary

event that all the a_i are negative has probability $1/2^n$, and so the probability that there is a positive input with positive output is $1 - 1/2^n$, which is indeed equal to $p(n+1, n)$.

Appendix: The recurrence relation for $r(n, d)$

This follows Wendel's [7] paraphrase of Schläfli's proof. The recurrence formula comes from analyzing how the number of regions changes as a hyperplane is added to the system. Let $H_1, \ldots, H_n$ be hyperplanes through the origin in $\mathbb{R}^d$. These hyperplanes are in general position, meaning that any intersection $k \leq d$ of them is a subspace of dimension $d - k$. Omit H_n for a moment and consider the regions created by the remaining $n - 1$ hyperplanes. There are $r(n-1, d)$ regions, which are of two types—those that meet H_n and those that don't. Let τ_1 and τ_2 be the number of each type; thus $r(n-1, d) = \tau_1 + \tau_2$. Now restore H_n to the system. It cuts each region of type 1 into two parts, and so $r(n, d) = 2\tau_1 + \tau_2$. Therefore, $r(n, d) = r(n-1, d) + \tau_1$.

Now τ_1 is also the number of regions in $H_n \cong \mathbb{R}^{d-1}$ created by the $n - 1$ hyperplanes $H_i \cap H_n$, $i = 1, \ldots, n-1$, and so $\tau_1 = r(n-1, d-1)$. Therefore, $r(n, d) = r(n-1, d) + r(n-1, d-1)$.

REFERENCES

1. T. M. Cover, The probability that a random game is unfair, *Ann. Math. Stat.* **37** (1966) 1796–1799. http://dx.doi.org/10.1214/aoms/1177699168
2. T. S. Ferguson, *Game Theory*, available at http://www.math.ucla.edu/~tom/Game_Theory/Contents.html.
3. G. Owen, *Game Theory*, 3rd ed., Academic Press, San Diego, 1995.
4. S. Roman, *Advanced Linear Algebra*, 3rd ed., Springer, New York, 2008.
5. L. Schläfli, *Theorie der vielfachen Kontinuität*, ed. J. H. Graf, Zürcher & Furrer, Zürich, 1901.
6. L. Schläfli, *Gesammelte mathematische Abhandlungen* Band 1, Birkhäuser, Basel, 1950.
7. J. G. Wendel, A problem in geometric probability, *Math. Scand.* **11** (1962) 109–111.

Summary A question in geometric probability about the location of the balls in a game of bocce leads to related questions about the probability that a system of linear equations has a positive solution and the probability that a random zero-sum game favors the row player. Under reasonable assumptions, we are able to find these probabilities.

KENT E. MORRISON is professor emeritus at California Polytechnic State University in San Luis Obispo, where he taught for thirty years. He received B.A. and Ph.D. degrees from the University of California, Santa Cruz. He has a number of research interests in the areas of algebra, geometry, probability, and combinatorics. Currently he is associated with the American Institute of Mathematics in Palo Alto, where he directs a project to encourage the use and development of open source and open access textbooks in mathematics.

The Use of Statistics in Experimental Physics

THOMAS J. PFAFF
Ithaca College
Ithaca, NY 14850
tpfaff@ithaca.edu

MAKSIM SIPOS
University of Illinois at Urbana-Champaign
Urbana, IL 61801
sipos2@illinois.edu

M. C. SULLIVAN
Ithaca College
Ithaca, NY 14850
mcsullivan@ithaca.edu

B. G. THOMPSON
Ithaca College
Ithaca, NY 14850
bthompso@ithaca.edu

MAX M. TRAN
Kingsborough Community College
Brooklyn, NY 11235
mtran@kingsborough.edu

When we teach statistics courses, the examples we use are typically slanted to the biological and health sciences or business and economics. We don't usually think of physics, even though statistics and probability are used extensively in this branch of science. For example, in statistical mechanics, physicists use statistical tools to analyze the behavior of systems with many degrees of freedom, and in quantum mechanics, physicists calculate probabilities of observing an outcome of a measurement. Uncertainty analysis and data analysis have long been tools of physics. Some examples can be quite complex, such as extracting random and systematic errors in an experiment from curve fits [**2**], while other examples, even from nuclear physics, are more straightforward [**3**].

Even straightforward tasks like residuals analysis can be quite rewarding. For example, examining the residuals from a Fourier analysis of tidal motion yields a more precise power spectrum [**7**]. As another example, comparing the measured flow of a viscous liquid to a model fit gives non-random residuals, which in turn suggests a way to improve the model to better describe the flow [**4**].

In this paper, we present in detail three examples of the use of statistics in experimental physics, any of which could be used as a case study in a statistics course. The first example involves a linear model of angular velocity. Here a residual plot shows that the linear model isn't the best model, and opens the door to a deeper understanding. In the second example, students see the use of residuals in a historically significant setting related to the quantum nature of radiation. This example was used in a course for introductory physics majors, as an exercise in data analysis using spreadsheets or Matlab. In the final example, goodness of fit is used to examine data from a nuclear decay experiment, one of the cornerstones of modern quantum physics and a triumph of 20th-century physics. In this experiment, the normal or Gaussian distribution is used to

Math. Mag. **86** (2013) 120–131. doi:10.4169/math.mag.86.2.120.

check the accuracy, or lack thereof, of new equipment intended for use in a junior-level laboratory course.

Drag on rotational motion

In this experiment a cylindrical aluminum disk is mounted on an air bearing, allowing it to spin in a horizontal plane. The air bearing pushes air under the disk, so that the disk is suspended by the air. This reduces frictional drag to a very low level, although some friction remains. The rushing air also imparts a certain turbine torque to the disk, so that if the disk is set in place and left for a long time, it will spin up until it reaches a terminal velocity. We denote the terminal velocity ω_∞, and measure it in units of radians per second (rad/s).

To begin the experiment, we spin the disk by hand to an initial angular velocity ω_0 (in the same direction as ω_∞ but faster). Then we measure the angular velocity $\omega(t)$ as a function of time t as the disc gradually slows down. The goal of the experiment is to specify an accurate mathematical model for the observed motion. We want to find a mathematical model such that only random fluctuations show in the difference between the data and the model—that is, so that the residuals have no pattern.

The details and graphs of this experiment are available in a separate publication [**12**].

Drag due to air friction is notoriously difficult to model theoretically. We choose to model the change in angular speed, $\omega'(t)$, as a function of $\omega(t)$ itself. It has been found that good fits can be obtained when this function is linear or quadratic. There is theoretical support for either a linear or quadratic model; alternatively, these models can be taken as Taylor approximations to a more general function.

The quadratic model can be written as

$$\omega' = c_T - c_1\omega - c_2\omega^2, \tag{1}$$

where c_T is the turbine torque and c_1 and c_2 are the linear and quadratic drag coefficients, all normalized to the disc's moment of inertia. Here, c_T, c_1, and c_2 all depend in some undetermined way on the pressure p in the air bearing (which can be varied in different trials of the experiment).

Under some conditions, the experiment can be modeled to first order, i.e., quadratic drag can be neglected. The above equation then reduces to

$$\omega' = c_T - c_1\omega,$$

and has a closed form solution

$$\omega - \omega_\infty = (\omega_0 - \omega_\infty)e^{-c_1 t}. \tag{2}$$

By taking a logarithm, equation (2) can be written in a linear form and a least-squares fit can be applied to obtain estimates of the parameters c_T and c_1.

The question is whether this model suffices under the conditions of the experiment. In fact, while it provides a good fit to the data, a plot of the residuals reveals a slight non-random discrepancy at early times. Early times correspond to higher angular speed, where the quadratic term is expected to have more of an effect and the second order equation may be a better model.

When the quadratic drag is present, (1) has the closed form solution

$$\omega = \frac{\omega_\infty + (\omega_\infty + c_1/c_2)g(\omega_0)e^{-st}}{1 - g(\omega_0)e^{-st}}, \tag{3}$$

where $s = \sqrt{c_1^2 + 4c_2c_T}$,

$$g(\omega) = (\omega - \omega_\infty)/(\omega + \omega_\infty + c_1/c_2),$$

and

$$\omega_\infty = (s - c_1)/2c_2.$$

The nonlinear form of this solution resists linearization. Since there are three co-dependent fitting parameters, c_T, c_1, and c_2, it is difficult (but not impossible) to manually fit the model to the experimental data. As an alternative, we can use a computer method (such as `fminsearch` in Matlab) to perform a nonlinear fitting and parameter estimation. The plot in FIGURE 1 shows the values obtained for c_2 for several different air bearing pressures that were obtained using an automated three-parameter fit. We found general agreement between the values obtained by fitting by hand and those obtained by using the automated method.

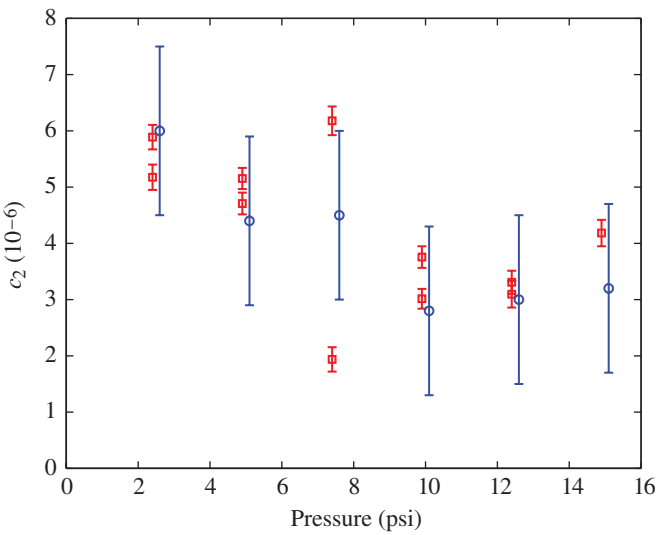


Figure 1 Values of c_2 computed from the fit of measured angular velocities to (3). The small error bars show errors computed by the numerical fitting procedure. The larger bars indicate errors estimated on the basis of unknowns in the experimental conditions.

In addition to the workload reduction that results from using an automated fitting method, it also gives estimates of the error in the fitted parameters. However, in this experiment, multiple runs of the experiment for the same values of the air bearing pressure, show that the estimated errors of c_2 calculated from the fit (smaller bars in FIGURE 1) are smaller than the actual errors estimated from experimental data (larger bars in FIGURE 1). Here, the additional error may be due to fluctuations in the air bearing pressure and in the measurement of ω. These sources of error were not reflected in the statistical model used in `fminsearch`, so, in this case, it is not justified to simply use the statistical error in determining the uncertainty in the value of c_2.

Overall, the experimental data are obtained easily and the simple linear model provides a good fit. Yet the slight non-random discrepancy between the linear model and the data shows up in a plot of the residuals, thus emphasizing their value in fitting models to experimental data. Once a quadratic model is used, the residuals no longer show a pattern. This demonstrates that linear and quadratic drag are simultaneously active in

the system. Further, this experiment demonstrates that we cannot always use the statistical quality-of-fit estimates to find error bars of experimental measurements. Instead, we need to keep in mind the global picture of the experiment, and what experimental conditions may make error bars larger.

The temperature of the universe

We will now use a theoretical thermometer to "take the temperature of the universe" by fitting a model, which has one parameter, to data. As in many such examples, we want to estimate the parameter—which is the temperature—and also assess how well the model fits the data. Both goals depend on analyzing residuals.

First, a brief overview of the physics involved. One of the crowning achievements of modern physics was the discovery of the quantum nature of radiation by Max Planck in 1900, including his explanation of the "blackbody" spectrum [**9**]. Another was the explanation of the origin of the universe in the "Big Bang"—a compact fireball, or singularity—together with evidence in the form of cosmic background radiation. Both of these achievements play roles in this exercise.

The blackbody spectrum Every object at a temperature above absolute zero emits electromagnetic radiation in a spectrum of frequencies, including light. This is called *thermal radiation.* An object that absorbs radiation perfectly (at all frequencies) is also a perfect emitter of thermal radiation. Such an ideal object is called a *blackbody* (since if it absorbs all light, it doesn't reflect it, and appears black at low temperatures). A good example of a blackbody in everyday life is a red-hot stove. It isn't black while it's radiating! Even so, most of the radiation the red-hot stove emits is not in the visible spectrum, but in the infrared, where its emissions are so intense that we can feel the radiation.

Planck was trying to understand the intensity-versus-frequency behavior of blackbody radiation—in other words, the spectrum of light from a hot body. The spectrum shows intensity of radiation as a function of frequency. By intensity we mean *absolute intensity*, reflecting corrections for distance and other factors. Expressions derived from classical physics were unable to explain the shape of the spectrum of a blackbody, as measured in the laboratory. By introducing the concept that electromagnetic radiation (light and radio waves) behaves as if it is composed of bundles of energy (photons), Planck derived a formula that does fit the data. This was the start of the quantum revolution in physics.

Planck's blackbody radiation equation is as follows:

$$I(f) = \frac{2hf^3}{c^2} \frac{1}{e^{hf/kT} - 1},$$

where I is the absolute intensity of the radiation ($W/m^2/Hz$) and f is the frequency (Hz). The symbols h, c, and k are all constants: h is Planck's constant (6.626069×10^{-34} J-s), c is the speed of light (2.99792458×10^8 m/s), k is the Boltzmann's constant (1.38065×10^{-23} J/K). That leaves T, the absolute temperature of the blackbody (Kelvin), as the only adjustable parameter in the equation. Thus, Planck's equation gives a unique spectrum (I as a function of f) for each value of the temperature T.

In FIGURE 2, we see a plot of I versus f for several different values of the temperature. We see that as the temperature increases, the peak intensity and peak frequency both increase.

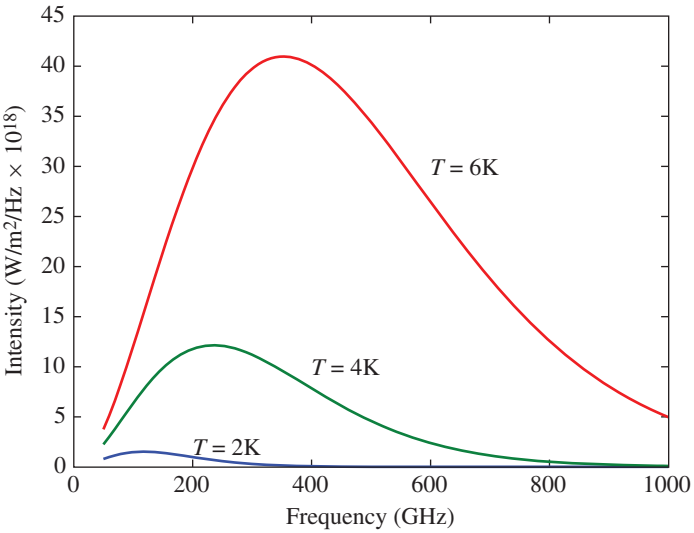


Figure 2 A plot of the intensity of the radiation of a blackbody versus frequency for temperatures of 2, 4, and 6 Kelvin.

Cosmic background radiation In 1964, Arno Penzias and Robert Wilson, two researchers at Bell Labs, discovered a source of noise in their radio antennas that they couldn't attribute to any earthly source [**8**]. Because the noise they measured came from all directions equally, there was no directed source for the radiation. This meant the radiation came from outside the solar system, and indeed, outside our galaxy!

Because the universe is so large, it behaves as a single object at a uniform temperature; in effect, the universe is a huge blackbody that began with the Big Bang and has been expanding and cooling ever since. In 1948, Alpher and Hermann [**1**] estimated on theoretical grounds that the current temperature of the universe is about 5 K. In talking with their colleague Robert Dicke, Penzias and Wilson realized that they were measuring the radiation emitted by the entire universe as described by Alpher and Hermann, and others. The noise they measured was the blackbody radiation left from the Big Bang, now called the cosmic microwave background radiation (CMB). Penzias and Wilson were able to relate the frequencies they measured to a blackbody of very low temperature, roughly 3 K.

In 1989, NASA launched the Cosmic Background Explorer (COBE) spacecraft, whose mission was to make detailed measurements of the CMB [**5**]. FIGURE 3 is an intensity versus frequency plot of data from COBE [**6**], together with a best-fit blackbody model using a temperature of 2.725 K. The uncertainties in the observed data points are not visible, as they are less than the width of the line. The remarkable agreement between these data and the model is obvious. More astounding is that this measurement of the universe as a whole can be characterized to a high degree of accuracy by a single number, the average temperature.

Estimating the parameter But what is the temperature, exactly? An accurate measurement is a key to understanding the Big Bang and the evolution of the universe. It is found by fitting a curve to the data, selecting T to minimize the residuals.

A plot of the residuals for the fit when $T = 2.725$ K, the middle row of stars in FIGURE 4, shows that the remaining discrepancies are random and are of a magnitude the same as the uncertainty in the data themselves. The root-mean-square deviation from the model fit is at the level of 50 ppm of the peak intensity. An informative

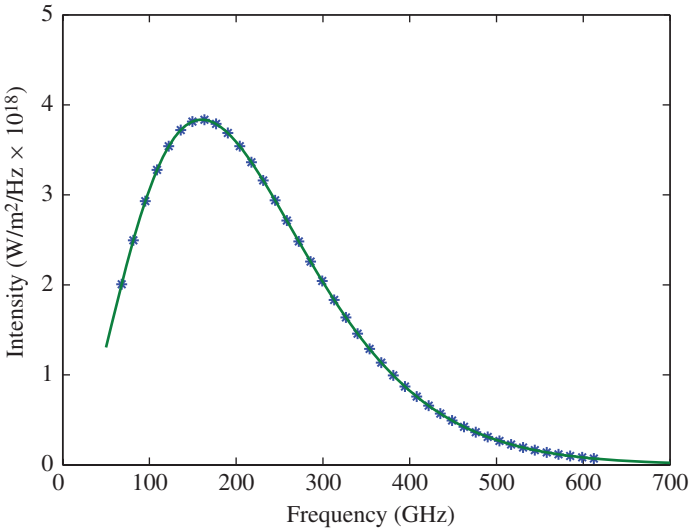


Figure 3 A plot of the microwave intensity versus frequency data from the COBE spacecraft (dots), together with a blackbody model with a temperature of 2.725 K (line). The uncertainty in the data measurements is approximately 2×10^{-22} W/m^2/Hz and thus is much smaller than the width of the line on the plot.

exercise is to change the model temperature and observe the change in the residual plot. Even a 0.001 K change—higher or lower—leads to residuals that show systematic discrepancies in the fit (FIGURE 4, the lower row of circles or the upper row of squares) even though the intensity plot is indistinguishable from the one shown in FIGURE 3.

We give our students the project of fitting these data with the blackbody model using trial and error, making sure that they produce a plot of the residuals in addition to the intensity plot and a calculation of the rms deviation. They see immediately how sensitive it is to small temperature changes, and are able to home in on the average temperature of the universe quickly.

Nuclear counting of cesium-137

Nuclear counting experiments are a staple of all instructional physics laboratories. These atomic decays occur spontaneously and randomly. In a common laboratory experiment, one of the neutrons inside ^{137}Cs spontaneously decays, becoming a proton, and ^{137}Cs becomes ^{137}Ba. When this decay occurs, the atom emits an electron, an antineutrino, and a high-energy gamma ray. Of the by-products of this nuclear decay, the gamma rays are the easiest to detect.

Students of mathematics are familiar with the fact that the rate of these spontaneous nuclear transformations is not constant and decreases with time. These transformations occur less and less frequently as time goes on, and follow an exponential decay. The decay constant is measured by the half-life; that is, the time required for half of the original material (in our case, ^{137}Cs) to decay into the product (^{137}Ba).

However, if the half-life is long, then on the time scale of an undergraduate laboratory experiment, the rate at which the nuclei spontaneously transform can be considered constant. Thus, studying the rate at which gamma rays are emitted from a sample of ^{137}Cs is an ideal example of a Poisson process in time. A Poisson process is an experiment in which events occur at random times that are uniform over an interval and independent of each other—which exactly describes nuclear decays. The process

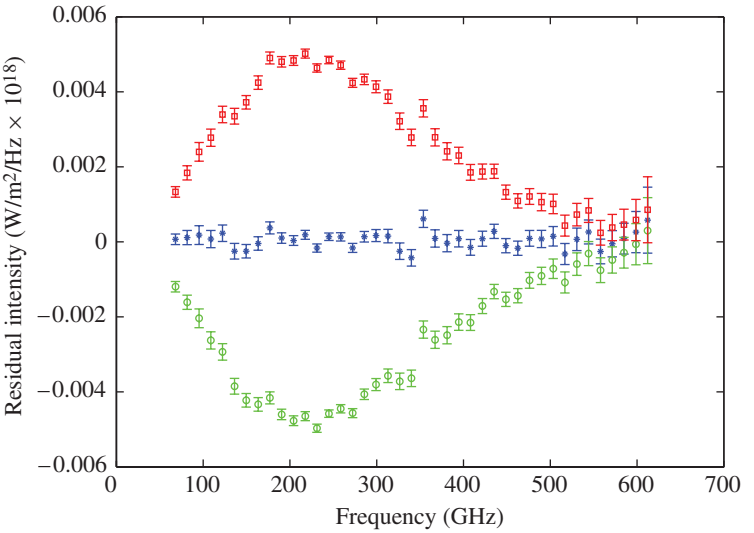


Figure 4 A plot of the residuals obtained by subtracting the model fit from the data in FIGURE 3. The uncertainties ($1\,\sigma$) in the intensities measured by COBE are also plotted and are of a magnitude of the model discrepancies. The fit for $T = 2.725$ K (middle row of stars) shows no systematic differences, just random ones. The rms deviation of the fit is 50 ppm of the peak intensity. However, a plot of the residuals obtained when the model temperature is changed by 0.001 K to 2.724 K (upper row of squares) or 2.726 K (lower row of circles) shows that there is a systematic difference between the model and data.

is characterized by one parameter, the rate, or the expected number of events per unit time. The rate has units of time^{-1}.

When used in a mathematics classroom, nuclear decay data can demonstrate the power of statistics for verifying physical data. A useful approach is to analyze subsets of the data, showing that for more and more data points, the fit to a bell curve improves. In what follows, nuclear decay can also be used as an example of the differences between Gaussian and Poisson distributions.

In the physical experiment, we place an ampule of ^{137}Cs a few centimeters from a scintillation detector. Every time a gamma ray is incident on the scintillating medium, the detector will convert the gamma ray into a voltage pulse with a peaked shape. This voltage pulse is then amplified and put through a linear gate and stretcher. The linear gate and stretcher converts the voltage pulses into square-wave pulses of given width (user-definable, between 0.5 and 5 μs). This process makes the pulses easier to detect with a computer. The square-wave output of the linear gate and stretcher is routed to the computerized data acquisition hardware. The goal is to convert each nuclear decay into a count registered by the computer. If this occurs, then we should expect a Poisson distribution of counts. If we do not see a Poisson distribution, then we have to search for the reason why.

To count the pulses, we have long used a microcomputer designed and built by one of our faculty members specifically for this experiment, nicknamed "Femto." In 2006, we purchased a new data acquisition card commercially available from National Instruments, the PCI-MIO-16E4, along with a simple Matlab routine to collect the data from the acquisition card. Our intent was to use the new card in place of Femto, but it had to be tested first.

To address the question of whether the National Instruments data acquisition system could replace Femto, we thought it wise to see whether the data acquisition card

could also accurately record the voltage pulses. Thus, we performed an experiment to compare our two data acquisition systems. The experiment consisted of placing a radioactive source a certain distance away from the scintillation detector. We then counted the number of decays in 16,400 periods of 0.1 s each, fitting both the Poisson distribution and the Gaussian distribution to the data.

In the experiment, we have n radioactive atoms, each with a certain probability p of decaying and being counted in a given time interval, and a probability of r counts during this time interval. Typically, we have n on the order of Avogadro's number, $n > 10^{25}$, $p \approx 10^{-20}$, giving an average number of decays $np \equiv \mu \approx 10^5$. However, the gamma rays emitted as a by-product of the radioactive decay are emitted in all directions. Our experimental detector can only detect gamma rays emitted directly toward our detector, so for these reasons, $r \approx 100$. Since n is large, we have $r \ll n$. This means that only a very small proportion of the atoms decay in each time-slice.

Under these circumstances there is no point in keeping track independently of n, p, or the fraction of decay events captured by the detector. The influence of all of these variables is encapsulated in the rate r. The distribution is closely approximated by the Poisson distribution with rate r, which is given by

$$p_{\text{Poisson}}(r) = \frac{(\mu)^r}{r!} e^{-\mu}. \tag{4}$$

We can approximate the true mean μ by the average value of r, called $\bar{r}$, which can be calculated from the data. Taking $\mu \approx \bar{r}$, we can fit the expected distribution $p_{\text{Poisson}}(r)$ to our data. We perform this fit setting the parameter $\mu = \bar{r}$.

When r, n, and $n - r$ are all large, the Poisson distribution approaches the Gaussian distribution, given by:

$$p_{\text{Gaussian}}(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r-\mu)^2/2\sigma^2} \tag{5}$$

$$= \frac{1}{\sqrt{2\pi np}} e^{-(r-\mu)^2/2np}. \tag{6}$$

It is a good exercise for math students to show that, for the Poisson distribution, the width of the distribution is the square root of the mean, therefore $\sigma = \sqrt{\mu}$. That is, while Gaussian distributions can arise from limits of Poisson distributions, not all Gaussians arise in this way—only those with $\sigma = \sqrt{\mu}$. This allows us to substitute $\sigma = \sqrt{np}$ in (5). If we again find $\bar{r}$, then we can take $\mu \equiv np \approx \bar{r}$ and fit our data using the Gaussian distribution. The Gaussian fit is performed by setting $\mu = \bar{r}$ and σ to the standard deviation calculated from the data. The Gaussian distribution has another useful prediction: In a Gaussian distribution, we expect 68% of the data to lie within $\mu \pm \sigma$ and 95.4% of the data to lie within $\mu \pm 2\sigma$.

Finally, we can test the goodness of our fits. A statistic often used to verify a quality of a fit is the chi-square (χ^2) test. For our problem, "χ^2 per degree of freedom" is defined as

$$\chi^2_\nu = \frac{1}{\nu} \sum_r \frac{\left(p_{\text{th}}(r) - p_{\text{exp}}(r)\right)^2}{p_{\text{th}}(r)},$$

where $p(r)$ is the theoretical or experimental probability distribution and ν is the number of degrees of freedom in the fit, equal to the number of different values of r minus the number of adjustable parameters in the fitting functions. For both of our datasets, $\nu \approx 100$. As the quality of the fit improves, we expect the numerator to become small; thus we expect the χ^2 statistic to become smaller with a better quality of fit.

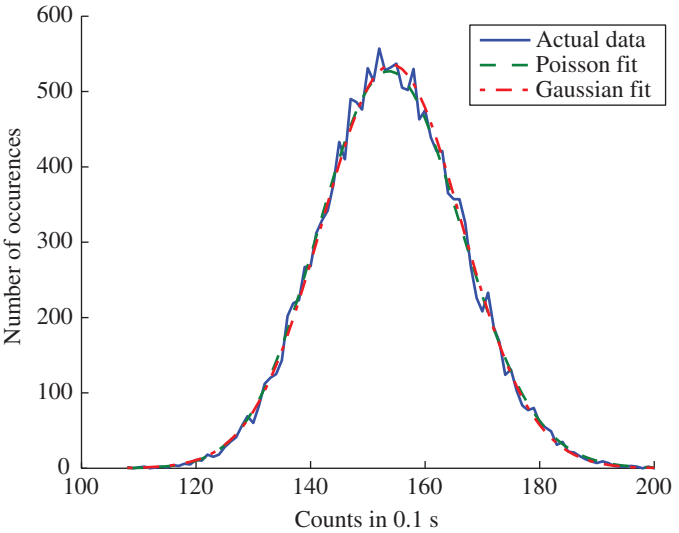


Figure 5 Histogram of the 16,400 decay count measurements using our home-built microcomputer "Femto." The mean was found to be $\bar{r} = 154.2 \pm 0.1$ per 0.1 s. The standard deviation of the distribution was found to be $\sigma = 12.2$ per 0.1 s. Gaussian (dot-dashed line) and Poisson (dashed) fits show excellent agreement between the data and the model. Analysis of the fits can be seen in TABLE 1.

The data from the experiment using the home-built microcomputer Femto is shown in FIGURE 5. In our experiment, n, r, and $n - r$ are large, so the radioactive decay data should fit the requirements of both the Poisson and the Gaussian distributions, and indeed, we see excellent agreement between the fits and the experimental data. The mean was found to be $\bar{r} = 154.2 \pm 0.1$ counts per 0.1 s, with a standard deviation of $\sigma = 12.2 \pm 0.1$ counts per $\sqrt{0.1}$ s. The goodness of fit as measured by χ^2 is similarly reasonable for both the Poisson and Gaussian fits. Moreover, we expect $\sigma = \sqrt{\bar{r}}$, and we find that $\sqrt{\bar{r}} = 12.4$ counts per $\sqrt{0.1}$ s, in good agreement with predictions. Finally, the data also obey what we expect from the Gaussian distribution: 69% of the data lie within $\mu \pm \sqrt{\bar{r}}$ (predicted: 68%), and 96% lie within $\mu \pm 2\sqrt{\bar{r}}$ (predicted: 95.4%). These results are collected in TABLE 1.

The data from the National Instruments card and Matlab, shown in FIGURE 6, is markedly different. We are unable to obtain good fits using the Poisson distribution. However, we can still get good fits (as measured by the goodness-of-fit parameter) for the Gaussian distribution. At this stage, a student might be tempted to throw away the Poisson fit and accept the results of the Gaussian fit. However, on closer inspection, we even see that the data do not behave as the good Gaussian fit predicts. As discussed above, counting experiments (such as radioactive decays), because they occur randomly but at a certain rate, should obey the Poisson distribution, and thus have a standard deviation where $\sigma = \sqrt{\bar{r}}$. The standard deviation in this experiment is 15.3 ± 0.1 per $\sqrt{0.1\text{s}}$, whereas $\sqrt{\bar{r}} = 11.8$, significantly smaller than we would expect. This carries through to the rest of the distribution: We find only 56% of the data lies between $\mu \pm \sqrt{\bar{r}}$ (predicted: 68%), and only 88% lies between $\mu \pm 2\sqrt{\bar{r}}$ (predicted: 95.4%).

This closer inspection indicates that there is additional error besides counting. Because the experiments were identical except for the computerized data acquisition systems, clearly the error must come from the National Instruments PCI-MIO-16E4 data acquisition card or the Matlab routine used to collect the data from the card. The additional error results in missed counts. These missed counts can be caused by two

TABLE 1: Comparison of the fits computed from the datasets for the National Instruments PCI-MIO-16E4 data acquisition card and the microcomputer Femto. The error bars for $\bar{r}$ are the standard deviation of the mean. σ is the standard deviation, and we expect $\sigma = \sqrt{\bar{r}}$. χ^2_P and χ^2_G are the values of the goodness-of-fit parameter for Poisson and Gaussian distribution fits, respectively. $\bar{r} \pm \sqrt{\bar{r}}$ and $\bar{r} \pm 2\sqrt{\bar{r}}$ represent the fraction of the data in the stated range. We get uniformly good fits using Femto, and a good Gaussian fit using the PCI-MIO-16E4 card, though less data is within $\bar{r} \pm 2\sqrt{\bar{r}}$ than is predicted from the Gaussian distribution.

	Microcomputer Femto	Nat. Inst. PCI-MIO-16E4 & Matlab
$\bar{r}$	154.2 ± 0.1 per 0.1 s	140.4 ± 0.1 per 0.1 s
σ	12.2 per $\sqrt{0.1\text{s}}$	15.3 per $\sqrt{0.1\text{s}}$
$\sqrt{\bar{r}}$	12.4 per $\sqrt{0.1\text{s}}$	11.8 per $\sqrt{0.1\text{s}}$
χ^2_P	0.85	64.7
χ^2_G	1.06	1.71
$\bar{r} \pm \sqrt{\bar{r}}$	69%	56%
$\bar{r} \pm 2\sqrt{\bar{r}}$	96%	88%

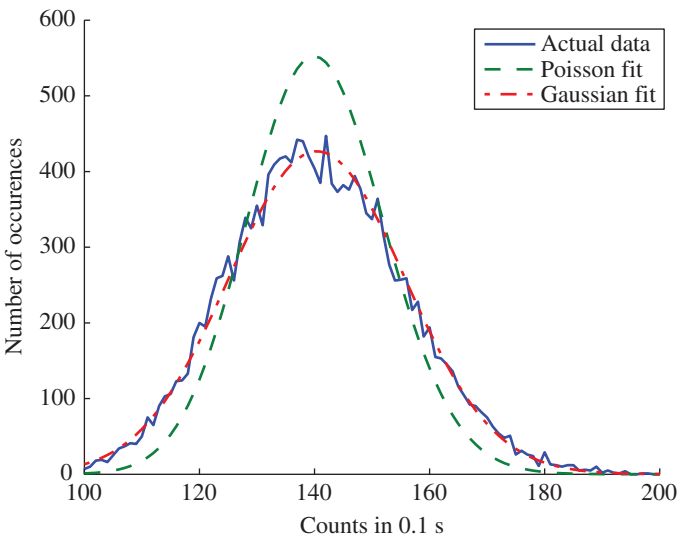


Figure 6 Histogram of the 16,400 decay count measurements using the National Instruments PCI-MIO-16E4 data acquisition card and a simple Matlab routine. The mean was found to be $\bar{r} = 140.4 \pm 0.1$. The standard deviation of the distribution was found to be 15.3 ± 0.1 per $\sqrt{0.1\text{ s}}$. The Gaussian (dot-dashed line) fit shows excellent agreement between the data and the model; however, the Poisson (dashed) fit shows poor agreement. Analysis of the fits can be seen in TABLE 1.

sources. They can occur when two pulses are too close to each other and are detected as one by the sampling hardware. Or, if pulses arrive during the time when the Matlab routine is converting analog voltage to a digital signal, those pulses are lost. The pulses counted had a width of 5 μs, which corresponds to a frequency of 200 KHz. The NI card's maximum sampling rate is 250 KHz, which is sufficient to count nearly all of the pulses. Thus, the counting error comes from the Matlab routine. Our simple

routine continually receives and converts voltages from analog to digital, resulting in many lost pulses when the computer is busy during the conversion. Because this error is random, and comes in addition to the intrinsic counting error, the result is a good fit of the Gaussian distribution but a poor fit of the Poisson distribution to the data, as seen in FIGURE 6. This also explains why we have recorded $\sigma \approx 15$ per $\sqrt{0.1\text{ s}}$ while $\sqrt{\bar{r}} \approx 12$ per $\sqrt{0.1\text{ s}}$. It is also clear that we are missing counts, as the experimental setup did not change but the average counts per 0.1 s decreased by approximately 14. Note that this error is somewhat fortuitous—had the error been systematic rather than random, neither the Poisson nor the Gaussian distribution would have fit the data with a reasonable goodness-of-fit parameter.

Thus we used two different computerized data collection systems to analyze the voltage pulses coming from a radioactive decay experiment. Using our home-built microcomputer, we find data that agree well with statistical predictions. Using a commercially available data acquisition system, we get good fits to the data using a Gaussian distribution, and poor fits using Poisson distribution. The take-home lesson here for mathematics students is to consider what a careful mathematical analysis of the data implies. Rather than accepting the Gaussian fit, a careful analysis of the errors in the fit leads us to understand what is wrong with using the commercial data acquisition system and even an estimate of how many counts it misses in every 0.1 s.

Conclusion

Statistical analysis plays a major role in several branches of physics, but it is also very useful for elementary physics. As the various experiments discussed in this paper illustrate, the use of statistical analysis, especially in the form of residual analysis, can help determine the best-fit model for experimental data. In other cases, it can be used to pinpoint flaws in laboratory equipment.

Our goal in writing this paper was to provide mathematicians with these examples. We do hope that everyone who read this enjoyed the connections between statistics and physics, but we also hope some of these examples find their way into statistics classrooms. This could happen in many ways. Instructors could simply take a few moments in class and discuss these examples, or even provide students with this paper. In other cases, ambitious instructors could use the examples here to create assignments for students. In the end, keep physics in mind the next time you are in a statistics classroom.

Acknowledgment The authors gratefully acknowledge Daniel A. Briotta Jr. for his amazing work and dedication to the task of undergraduate education displayed in his construction of the microcomputer "Femto."

REFERENCES

1. R. A. Alpher and R. C. Herman, Evolution of the universe, *Nature* **162** (1948) 680–682. http://dx.doi.org/10.1038/162680a0
2. J. Bechhoefer, Curve fits in the presence of random and systematic error, *Am. J. Phys.* **68** (2000) 424–429. http://dx.doi.org/10.1119/1.19457
3. Byron Curry, Dave Riggins, and P. B. Siegel, Data analysis in the undergraduate nuclear laboratory, *Am. J. Phys.* **63** (1995) 71. http://dx.doi.org/10.1119/1.17771
4. Rafael M. Digilov and M. Reiner, Weight-controlled capillary viscometer, *Am. J. Phys.* **73** (2005) 1020–1022. http://dx.doi.org/10.1119/1.2060718
5. D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, The Cosmic Microwave Background spectrum from the full COBE FIRAS data set, *Astrophysical Journal* **473** (1996) 576. http://dx.doi.org/10.1086/178173

6. Legacy Archive for Microwave Background Data Analysis. http://lambda.gsfc.nasa.gov/product/cobe/firas_monopole_get.cfm
7. Robert de Levie, Tidal analysis on a spreadsheet, *Am. J. Phys.* **72** (2004) 644–651. http://dx.doi.org/10.1119/1.1629086
8. A. A. Penzias and R. W. Wilson, A measurement of the excess antenna temperature at 4080Mc/s, *Astrophysical Journal* **142** (1965) 419–421. http://dx.doi.org/10.1086/148307
9. Max Planck, On the Law of Distribution of Energy in the Normal Spectrum, *Annalen der Physik* **4** (1901) 553. http://dx.doi.org/10.1002/andp.19013090310
10. S. H. Suyu, P. J. Marshall, M. W. Auger, S. Hilbert, R. D. Blandford, L. V. E. Koopmans, C. D. Fassnacht, and T. Treu, Dissecting the Gravitational Lens B1608+656, *Astrophysical Journal* **711** (2010) 201–222. http://dx.doi.org/10.1088/0004-637X/711/1/201
11. J. R. Taylor, *An Introduction to Error Analysis*, 2nd ed., University Science Books, Mill Valley, CA, 1997.
12. B. G. Thompson and P. A. Smith, An experiment in rotational motion with linear and quadratic drag, *Am. J. Phys.* **72**(4) (2004) 839–842. http://dx.doi.org/10.1119/1.1632493

Summary Most mathematicians are aware of the importance of statistics in biological sciences, business, and economics, but are less aware that statistics is used every day in experimental physics. This paper gives three interesting examples of how statistics plays a vital role in physics. These examples use the basic statistical tools of residuals analysis and goodness of fit.

THOMAS J. PFAFF is an associate professor of mathematics at Ithaca College. His main teaching interest is incorporating sustainability and quantitative literacy issues into mathematics courses. Pfaff is the lead-PI of the Multidisciplinary Sustainability Education Project, and he has given several presentations on sustainability in the mathematics curriculum. He also maintains a page of sustainability curriculum materials for mathematics at http://www.ithaca.edu/tpfaff/sustainability.htm.

MATTHEW C. SULLIVAN is an associate professor of physics at Ithaca College. His research focuses on the growth and properties of high-temperature superconductors. He is best known for his work on superconducting demonstrations using superconductors and magnetic tracks to demonstrate "quantum levitation" and "quantum locking."

MAKSIM SIPOS received bachelors degrees in physics and mathematics from Ithaca College and is now doing postgraduate studies in physics at the University of Illinois at Urbana-Champaign. His research interests are in bioinformatics, microbial ecology, and statistical mechanics.

BRUCE THOMPSON has degrees from two of the many CU's, the University of Colorado and Cornell University. He spent several years traveling to hot spots of the world to help control oil and gas well blowouts. Currently, he is a faculty member of the physics department at Ithaca College with interests in finite element time domain modeling of electromagnetic fields, seismic signals produced by elephants, and magneto-optical traps.

MAX TRAN is currently an assistant professor of mathematics and computer science at Kingsborough Community College. He has lifelong interests in the sciences and mathematics, and is currently engaging in research in physics, biology, and mathematics, focusing on the use of geometric algebra to encode and solve electromagnetic equations.

NOTES

A Continuous Function That Is Differentiable Only at the Rationals

MARK LYNCH
Millsaps College
Jackson, MS 39210
lynchmj@millsaps.edu

This note was motivated by two questions asked in an internet mathematics forum: Does there exist a continuous function that is differentiable only at the irrationals, and does there exist one that is differentiable only at the rationals? This question seems similar to the question: Is there a function that is continuous only at the irrationals and another that is continuous only at the rationals? Since the answers in the case of continuous functions are "yes" and "no," respectively, the author assumed that the answers for differentiable would again be "yes" and "no," respectively.

However, both answers are "yes." A continuous function differentiable only at the irrationals can be obtained by taking the antiderivative of the function given by Rudin [**5**, Remark 4.31]. Continuous functions that are differentiable only at the rationals are shown to exist by Zahorski [**6**], but without an explicit construction.

In this note, we give an explicit construction of a continuous function on $[0, 1]$ that is differentiable only at the rationals in $(0, 1)$. We do so by generalizing the technique used in [**2**].

The rhomboid

Let n be a positive integer. Let $\ell(x) = rx + s$ be a linear function defined on an interval $[a, b]$ with slope strictly exceeding n in absolute value. Then, its graph L is a line segment in the plane. We will restrict ourselves to the upward-sloping case $r > n$, since the case of $r < -n$ is similar.

Let $\epsilon > 0$ and consider the linear functions $\ell(x) + \epsilon$ and $\ell(x) - \epsilon$, with graphs L_1 and L_2 respectively, as in FIGURE 1. Let $m = (a + b)/2$. Then we can choose ϵ small enough so that $\epsilon < 1/n$, and also

$$\left| \frac{((\ell(b) - \epsilon) - (\ell(m) + \epsilon)}{(b - m)} \right| > n.$$

The rhomboid R is the closed region bounded by the lines L_1 and L_2 and by the vertical lines $x = a$ and $x = b$. The lines L, L_1, and L_2 all have slope greater than n, and we have chosen ϵ small enough so that the slope of the dotted line in FIGURE 1 is greater than n as well. As shorthand, we refer to R as a "rhomboid with slope n."

If the graph of a continuous function is contained in R, then it must have many chords of slope greater than n; in fact, every point on the graph is an endpoint of such a chord. We state this as a lemma.

Math. Mag. **86** (2013) 132–135. doi:10.4169/math.mag.86.2.132.

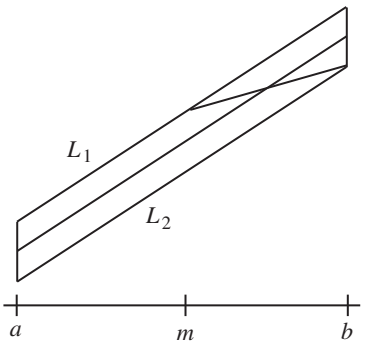


Figure 1 The rhomboid

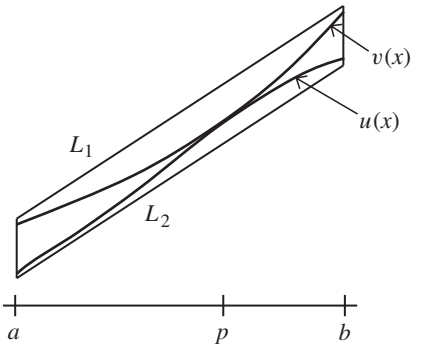


Figure 2 The pinched set

LEMMA 1. *If f is a continuous function on $[a, b]$ whose graph is contained in R, then for each $x \in [a, b]$, there is a $y \in [a, b]$ such that $|f(y) - f(x)|/|y - x| > n$.*

Proof. Assume that we are dealing with the case $r > n$ as in FIGURE 1. Let $x \in [a, b]$ and assume that $x \leq m$. Then we may choose $y = b$. Then,

$$\frac{|f(b) - f(x)|}{|b - x|} \geq \frac{|(\ell(b) + \epsilon) - (\ell(x) - \epsilon)|}{|b - x|} \geq \frac{|(\ell(b) + \epsilon) - (\ell(m) - \epsilon)|}{|b - m|} > n,$$

as required. For $x \geq m$, we take $y = a$ with a similar outcome. ∎

The pinched set

Let R be the rhomboid about L as described above. We can "pinch" this rhomboid at a point $p \in (a, b)$ as follows: Let u and v be two differentiable functions on $[a, b]$ whose graphs are contained in R and such that:

(a) $u(a) < \ell(a) < v(a)$, $u(b) < \ell(b) < v(b)$, and $u(x) < v(x)$ for $x \neq p$;
(b) $u(p) = v(p)$ and $u'(p) = v'(p)$.

See FIGURE 2. The pinched set P is the closed region between the graphs of u and v, and we say that P is pinched at p.

LEMMA 2. *If f is continuous on $[a, b]$ and the graph of f is a subset of P, then $f'(p) = u'(p) = v'(p)$.*

Proof. This follows from the squeeze principle. For example, if $p < x \leq b$, then

$$\left|\frac{u(x) - u(p)}{x - p}\right| \leq \left|\frac{f(x) - f(p)}{x - p}\right| \leq \left|\frac{v(x) - v(p)}{x - p}\right|,$$

and letting x converge to p, we see that the right-hand derivative of f at p equals $u'(p)$. ∎

The function

Let $\{q_1, q_2, q_3, \ldots\}$ be an enumeration of the rationals in $(0, 1)$. Let g_1 be a piecewise linear function on $[0, 1]$ such that the absolute value of the slope of each segment exceeds 1, and the segment boundaries occur at irrational points; see FIGURE 3(a).

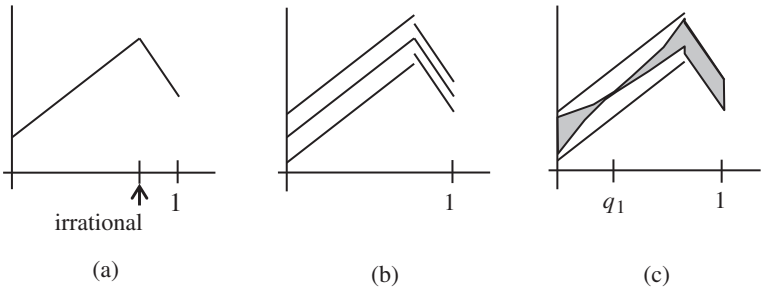


Figure 3 The function

Construct rhomboids about each segment of g_1 using $n = 1$, as in FIGURE 3(b). One of the rhomboids contains the rational number q_1 in its domain. Pinch that rhomboid at the point q_1, as in FIGURE 3(c). We denote the pinched set by P_1.

Let Z_1 be the union of the non-pinched rhomboid(s) and the pinched set P_1.

As the first step in an induction, let g_2 be a piecewise linear function on [0, 1] whose graph is contained in the interior of Z_1, such that the absolute value of the slope of each segment exceeds 2 and the segment boundaries occur at irrational numbers. Because of the pinched set in Z_1, the graph of g_2 necessarily consists of infinitely many segments. We attach the pinched point of Z_1 to the graph of g_2 so that g_2 is defined and continuous on [0, 1].

Construct rhomboids of slope 2 about each line segment of g_2 in the interior of Z_1. One of these rhomboids contains q_2 in its domain; let P_2 be the pinched set obtained by pinching that rhomboid at the point q_2. Let Z_2 be the union of P_2 and all the non-pinched rhomboids, together with the pinched point of Z_1 at q_1.

Continue the induction. In general, Z_n is a closed subset of Z_{n-1} that is the union of a pinched set P_n pinched at q_n, plus the rhomboids of slope n about the other line segments of a piecewise linear function g_n on [0, 1], together with pinched points at $q_1, \ldots, q_{n-1}$. Thus $\{Z_n\}$ is a nested sequence of closed and bounded sets above the interval [0, 1]. Moreover, each rhomboid of Z_n satisfies Lemma 1, and each pinched set P_n satisfies Lemma 2 at q_n.

Let $Z = \cap Z_n$. Then Z is a closed and bounded set, and its domain is [0, 1] because for each $x \in [0, 1]$ the sets $\{y \in [0, 1] : (x, y) \in Z_n\}$ are closed, bounded and nonempty, so their intersection $\{y \in [0, 1] : (x, y) \in Z\}$ is also nonempty. In constructing the rhomboids of slope n, we chose $\epsilon < 1/n$, forcing the sequence of ϵ values to approach zero, so this intersection has exactly one point. Therefore Z is the graph of a function f on [0, 1]. Since its graph is closed and bounded, f is continuous [**4**, Example 6, p. 169]. By Lemma 2, f is differentiable at each pinch point, so it is differentiable at the rationals.

THEOREM. *f is not differentiable at the irrationals.*

Proof. Consider irrational $x \in [0, 1]$, $\delta > 0$, and $M > 0$. Note that $x \notin P_n$ for infinitely many n; for example, given $x < \frac{1}{2}$, infinitely many P_n are subsets of $[\frac{1}{2}, 1]$. Choose n so large that $n \geq M$, $x \notin P_n$, and the domain of each rhomboid in Z_n has length less than δ. Let R be the rhomboid in Z_n with x in its domain. By Lemma 1, there exists y in the domain of R such that $\left|\frac{f(y)-f(x)}{y-x}\right| > n \geq M$; and of course $|x - q| < \delta$. Therefore

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x}$$

cannot be a finite number and f is not differentiable at x. ∎

We finish with some remarks.

(a) We can pinch the rhomboids at each stage to give the derivative $f'(q)$ whatever value we want at each rational point q. For example, we can arrange for $f'(q) = 0$ for all rationals q.

(b) If we don't pinch any rhomboids, then f is a continuous nowhere-differentiable function.

(c) There is a limit to the flexibility we have in choosing the behavior of f at the irrationals. For example, the result and proof above leave open this possibility: Can it be arranged that $f'(x) = \infty$ for irrationals x, and $f'(q) = 1$ for all rationals q? We show that this is not possible. Such a continuous function would be a strictly increasing function as noted in [**3**], where the hypothesis $f'(x) = \infty$ is explicitly allowed. Thus f would have a continuous inverse g and we would have $g'(f(x)) = 0$ for irrational x. By [**1**, item (18.41)(d)], g would be absolutely continuous and $g(x) - g(0) = \int_0^x g'(t)\,dt = 0$ for all $x \in [0, 1]$, a contradiction.

REFERENCES

1. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.
2. M. Lynch, A continuous, nowhere differentiable function, *Amer. Math. Monthly* **99** (1992) 8–9. http://dx.doi.org/10.2307/2324541
3. A. D. Miller and R. Výborný, Some remarks on functions with one-sided derivatives, *Amer. Math. Monthly* **93** (1986) 471–475. http://dx.doi.org/10.2307/2323476
4. K. Ross, *Elementary Analysis: The Theory of Calculus,* 2nd ed., Springer-Verlag, New York, 2012.
5. W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 1964.
6. Z. Zahorski, Sur l'ensemble des points de non-derivabilite d'une function continue, *Bull. Soc. Math. France* **74** (1946) 147–178.

Summary An explicit construction is given of a function that is continuous on an interval, and differentiable only at the rationals.

Geometry of Cubic Polynomials

SAM NORTHSHIELD
SUNY, Plattsburgh, NY 12901
northssw@plattsburgh.edu

Imagine a sphere with its equator inscribed in an equilateral triangle. This Saturn-like figure will help us understand from where Cardano's formula for finding the roots of a cubic polynomial $p(z)$ comes. It will also help us find a new proof of Marden's theorem, the surprising result that the roots of the derivative $p'(z)$ are the foci of the ellipse inscribed in and tangent to the midpoints of the triangle determined by the roots of the polynomial.

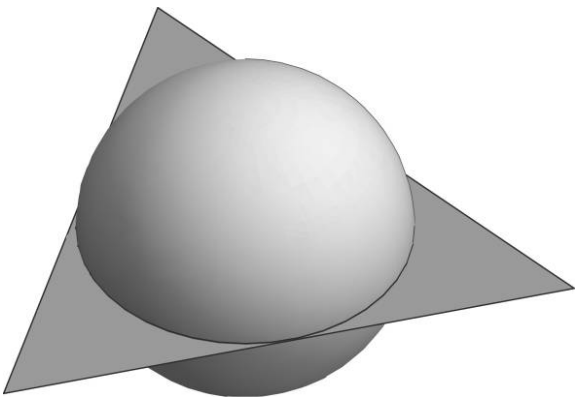


Figure 1 An equilateral triangle and its inscribed sphere

Marden's theorem and its real version

By the fundamental theorem of algebra, a cubic polynomial, with real or complex coefficients, has precisely three (possibly complex) roots. In general, these roots are distinct and don't all lie on a line. That is, they form a triangle in the complex plane. The Gauss–Lucas theorem [**3**] implies that the two roots of the derivative must lie inside this triangle. It states that for any polynomial $p(x)$ with complex coefficients, the roots of $p'(x)$ all lie within the complex hull of the roots of $p(x)$. We present an elegant proof for the cubic case. If r, s, t are the (complex) roots of $p(x)$, and if $p'(u) = 0$ where u is *not* in the complex hull of $\{r, s, t\}$, then there exists θ such that each of $e^{i\theta}(u - r)$, $e^{i\theta}(u - s)$, and $e^{i\theta}(u - t)$ has positive real part. This, however, would contradict the fact that

$$0 = \overline{p'(u)/p(u)} = \frac{u - r}{|u - r|^2} + \frac{u - s}{|u - s|^2} + \frac{u - t}{|u - t|^2}.$$

The proof extends easily to higher dimensions.

This paper arose from an attempt to understand Marden's theorem, the subject of a fascinating article by D. Kalman [**4**]. This theorem has a long history that goes back at

Math. Mag. **86** (2013) 136–143. doi:10.4169/math.mag.86.2.136.

least to work by Siebeck in 1864—see [**4**] for the history of the theorem and some of its generalizations. The theorem states that for a cubic polynomial $p(z)$ whose roots form a triangle in the complex plane, the roots of $p'(z)$ are the foci of the (unique) ellipse inscribed in that triangle tangent to the midpoints of the sides. FIGURE 2 illustrates Marden's theorem. This is an analog of the fact that the midpoint of the two roots of a quadratic polynomial $p(x)$ coincides with the root of its derivative $p'(x)$.

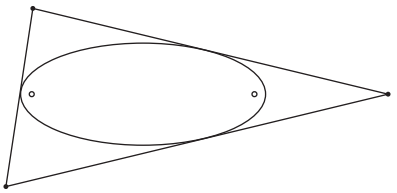


Figure 2 Roots of p' are foci of "midpoint" ellipse.

Note that FIGURE 2 does not make much sense if the roots of $p(x)$ are collinear, since the ellipse disappears. What is going on this case? In order to draw a sphere, we can draw a circle with an ellipse inside (FIGURE 1). That ellipse is a projection (onto the plane) of the equator of the sphere. This illustrates a technical fact: The projection of a circle is an ellipse. We shall, in due time, prove this (and provide a definition of "ellipse").

Consider an equilateral triangle in 3-space, with a sphere such that one of the sphere's great circles is inscribed in the triangle. FIGURE 1 shows a projection of this figure. A sufficient rotation of the sphere-with-triangle figure in 3-space will yield, to our vantage point, the disappearance of the triangle and the ellipse, but not of the sphere. The foci of the now-flattened ellipse become the ends of one of the diameters of the sphere, and we may deduce the following result (with formal proof supplied below).

THEOREM. *Any three real numbers, not all equal, are the projections of the vertices of some equilateral triangle in the plane. For a cubic polynomial $p(x)$ with three real roots (not all equal), the inscribed circle of the equilateral triangle that projects onto those roots itself projects to an interval with endpoints equal to the roots of $p'(x)$.*

Proof. Suppose we have a polynomial $p(x)$ with three real roots r, s, t. That is, we may take

$$p(x) = (x-r)(x-s)(x-t) = x^3 - (r+s+t)x^2 + (rs+rt+st)x - rst.$$

It is easy to verify that the points in $\mathbb{R}^2$ with coordinates $(r, (s-t)/\sqrt{3})$, $(s, (t-r)/\sqrt{3})$, and $(t, (r-s)/\sqrt{3})$ form the vertices of an equilateral triangle whose first coordinates coincide with the roots of $p(x)$ (since the distance between any two is the symmetric function $2\sqrt{(r^2+s^2+t^2-rs-rt-st)/3}$). Note that the inscribed circle has center $((r+s+t)/3, 0)$ and radius $1/\sqrt{12}$ times the distance between any two of the vertices (i.e., the radius is $\frac{1}{3}\sqrt{(r^2+s^2+t^2-rs-rt-st)}$). Hence, the first coordinate projection of the inscribed circle is the closed interval $\frac{1}{3}\big(r+s+t \pm \sqrt{(r^2+s^2+t^2-rs-rt-st)}\big)$. By the quadratic formula, the endpoints are readily seen to be the roots of $p'(x) = 3x^2 - 2(r+s+t)x + (rs+rt+st)$. ∎

FIGURE 3 shows the graph of a cubic polynomial and its relation to an equilateral triangle and its inscribed circle. This includes, after all, another view of the Saturn-like figure.

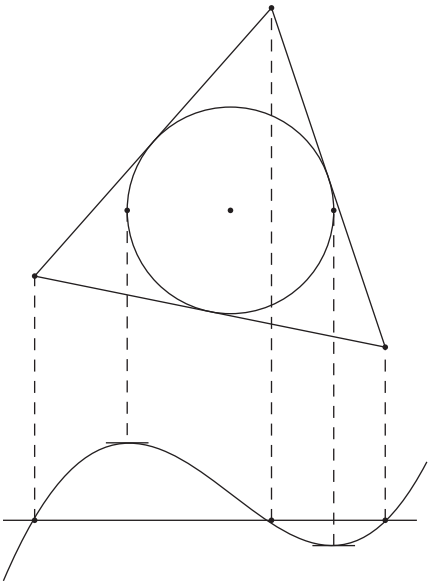


Figure 3 A cubic with corresponding triangle and circle

This reminds us of Hadamard's dictum: *The shortest path between two truths in the real domain passes through the complex domain.*

Ellipses and linear maps

We recall the familiar way of drawing an ellipse by hand; given two points (possibly coincident) and a piece of string connecting them with enough length to have some slack, moving a pencil around while keeping the string taut results in an ellipse (see FIGURE 4). That is, the set of all points

$$\{x \in \mathbb{C} : |x - u| + |x - v| = L\}$$

where $u, v \in \mathbb{C}$ and $L \in (|u - v|, \infty)$ is an ellipse. We may take this as a definition—i.e., every ellipse arises this way—since any ellipse has a unique set of foci $\{u, v\}$ and a maximum width L such that the previous construction yields the ellipse [**9**].

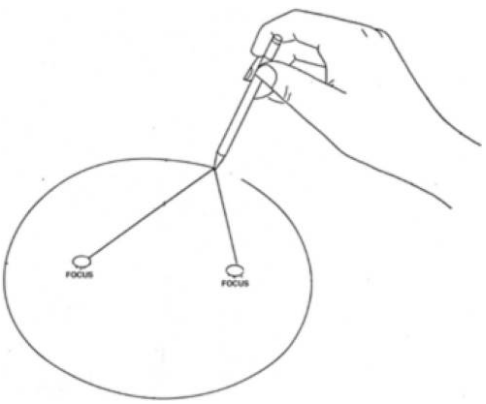


Figure 4 Drawing an ellipse

We stated earlier that, generally, a projection of a circle in $\mathbb{R}^3$ onto a plane is an ellipse. This fact is tantamount to saying that the image of a circle $\mathbb{R}^2$ under a linear map is an ellipse (since rotations are linear). We shall prove this.

Recall that linear map from $\mathbb{R}^2$ to itself is of the form

$$(x, y) \mapsto (\alpha x + \beta y, \gamma x + \delta y)$$

for some $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Thinking of this map as a map instead from $\mathbb{C}$ to itself, i.e., $z := x + iy$ maps to

$$\begin{aligned}&\alpha x + \beta y + i(\gamma x + \delta y)\\ &\quad = \frac{1}{2}(\alpha + \delta + i(\gamma - \beta))(x + iy) + \frac{1}{2}(\alpha - \delta + i(\gamma + \beta))(x - iy),\end{aligned}$$

we see that every linear map is of the form

$$z \mapsto az + b\overline{z}$$

for some complex a, b.

LEMMA. *Every one-to-one linear map $az + b\overline{z}$ takes the unit circle to an ellipse with foci $\pm 2\sqrt{ab}$.*

Proof. The unit circle can be parameterized by

$$C := \{e^{i\theta} : 0 \le \theta < 2\pi\},$$

and so if E is the image of the circle under the linear map $az + b\overline{z}$, then any $x \in E$ may be written $x = ae^{i\theta} + be^{-i\theta}$ for some θ. Let

$$z := \sqrt{a}\, e^{i\theta/2} \quad \text{and} \quad w := \sqrt{b}\, e^{-i\theta/2}.$$

Then

$$\begin{aligned}\left|x - 2\sqrt{ab}\right| + \left|x + 2\sqrt{ab}\right| &= \left|ae^{i\theta} + be^{-i\theta} - 2\sqrt{ab}\right| + \left|ae^{i\theta} + be^{-i\theta} + 2\sqrt{ab}\right|\\ &= |z - w|^2 + |z + w|^2 = 2|z|^2 + 2|w|^2 = 2|a| + 2|b|\end{aligned}$$

and thus E is an ellipse with foci $\pm 2\sqrt{ab}$. ∎

Marden and Cardano

Suppose now that $p(z)$ is a cubic with complex roots r, s, t. For convenience, we shall assume $r + s + t = 0$ so that we may assume

$$p(z) = z^3 + (rs + rt + st)z - rst.$$

Just as any three real numbers, not all equal, are the projections of the vertices of an equilateral triangle, it is also clear that every three complex numbers, not all equal, are the projections of the vertices of an equilateral triangle in 3-space. Considering this triangle and its inscribed sphere as in FIGURE 1, there should be a linear map from the plane to itself taking the unit circle to an ellipse containing r, s, t. FIGURE 5 illustrates how such a map would take the equilateral triangle with vertices corresponding to third roots of unity $1, \omega, \overline{\omega}$ (along with inscribed and circumscribed circles) to the triangle with vertices r, s, t (along with inscribed and circumscribed ellipses). Since linear maps take midpoints to midpoints (i.e., $f((z + w)/2) = (f(z) + f(w))/2$), the

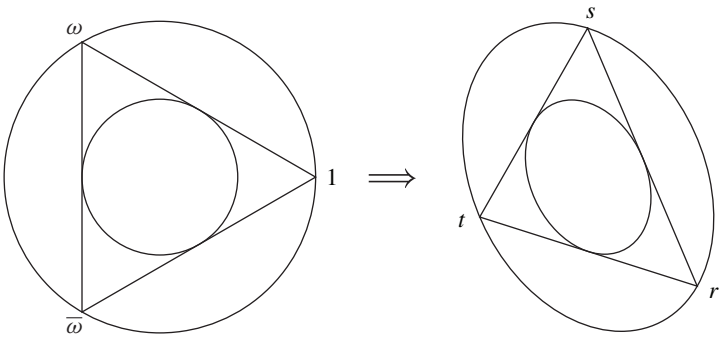


Figure 5 $f(z) = az + b\overline{z}$, where $f(1) = r$ and $f(\omega) = s$

inscribed ellipse is tangent to the triangle at midpoints of the sides and, further, since the outer circle has radius twice that of the inner one, the outer ellipse is a dilation of the inner one by a factor of 2.

Of course, we may find a, b explicitly by solving the simultaneous equations $a + b = r$ and $a\omega + b\overline{\omega} = s$ (where, of course, $\omega = (-1 + i\sqrt{3})/2$). Because $r + s + t = 0$, it follows that $a\overline{\omega} + b\omega = t$.

Using this, we see that $rs + rt + st = -3ab$ and $rst = a^3 + b^3$, and thus we may rewrite $p(z)$ in terms of a and b:

$$p(z) := z^3 - 3abz - (a^3 + b^3). \tag{1}$$

Alternatively, if $w := ae^{i\theta} + be^{-i\theta}$, then it is easy to see that

$$w^3 - 3abw = a^3e^{i3\theta} + b^3e^{-i3\theta}$$

and so the roots of $z^3 - 3abz - (a^3 + b^3)$ are $\{ae^{ik\pi/3} + be^{-ik\pi/3} : k = 0, 2, 4\}$ and equation (1) follows.

As still another alternative (see [**5**]), it is easy to verify

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \overline{\omega} \\ 1 & \overline{\omega} & \omega \end{pmatrix} \cdot \underbrace{\begin{pmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & t \end{pmatrix}}_{D} = \underbrace{\begin{pmatrix} 0 & a & b \\ b & 0 & a \\ a & b & 0 \end{pmatrix}}_{M} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \overline{\omega} \\ 1 & \overline{\omega} & \omega \end{pmatrix}.$$

So the matrices M and D have the same characteristic polynomials:

$$p(z) = \det(zI - M) = z^3 - 3abz - (a^3 + b^3).$$

MARDEN'S THEOREM. *If $p(z)$ is a cubic polynomial with three complex roots r, s, t that form a triangle in $\mathbb{C}$, then the roots of $p'(z)$ are the foci of the unique ellipse tangent to the midpoints of each side.*

Proof. By the Lemma, the foci of the outer ellipse in FIGURE 5 are $\pm 2\sqrt{ab}$, and so the inner ellipse has foci $\pm\sqrt{ab}$ that, by equation (1), are the roots of $p'(z)$. ∎

Equation (1) also provides a method for solving cubic equations. That is, if a, b can be found so that a given (reduced) cubic is of the form $z^3 - 3abz - (a^3 + b^3) = 0$, then its roots are $a + b$, $a\omega + b\overline{\omega}$, and $a\overline{\omega} + b\omega$. The result is an old one due to Cardano and the method of proof also goes back, apparently, to Lagrange. Among the many references to Cardano's formula are a web version [**10**] and a geometric approach similar to ours [**6**]. See also Chapter 6 of [**1**].

CARDANO'S FORMULA. *The solutions of the equation*

$$z^3 - 3Az - 2B = 0$$

are $a + b$, $a\omega + b\overline{\omega}$, *and* $a\overline{\omega} + b\omega$, *where*

$$a, b := \sqrt[3]{B \pm \sqrt{B^2 - A^3}}.$$

Proof. Starting with $z^3 - 3Az - 2B$, we seek a, b such that $ab = A$ and $(a^3 + b^3)/2 = B$. Given such a, b, note that

$$0 = (z - a^3)(z - b^3) = z^2 - (a^3 + b^3)z + a^3b^3 = z^2 - 2Bz + A^3;$$

so

$$a^3, b^3 = B \pm \sqrt{B^2 - A^3},$$

and the result follows. ∎

Although every real number has a unique real cube root (the usual meaning for $\sqrt[3]{\ }$), the cube root of a complex number is not uniquely defined. In the formula above, we may choose a, b to be cube roots of $B + \sqrt{B^2 - A^3}$ and $B - \sqrt{B^2 - A^3}$, respectively, such that $ab = A$.

EXAMPLE. $f(x) = x^3 - 3x^2 - 12x + 18$.

With $f(x)$ as above, let $p(x) := f(x + 1) = x^3 - 15x + 4$. In general, if $p(x) = f(x - a/3)$, where a is the coefficient for x^2 in f (and f is monic), then the x^2 coefficient of p is 0. A cubic with zero x^2 coefficient is sometimes called a "depressed cubic." To equate with $x^3 - 3Ax - 2B$, we let $A = 5$ and $B = -2$. Then $B^2 - A^3 = -121$ and

$$a^3, b^3 = -2 \pm 11i.$$

Although a^3 has three cube roots, finding one gives the rest if we multiply by ω and $\overline{\omega}$. The method of finding the cube roots of a given complex number is quite elaborate, but it is not hard to verify that $(-2 + i)^3 = -2 + 11i$, and so we may take $a := -2 + i$. By the condition $ab = A = 5$, we must then take $b := -2 - i$. It follows that the roots of $p(z)$ are

$$\{a + b, a\omega + b\overline{\omega}, a\overline{\omega} + b\omega\} = \{-4, 2 + \sqrt{3}, 2 - \sqrt{3}\},$$

and thus the roots of $f(z)$ are -3, $3 + \sqrt{3}$, and $3 - \sqrt{3}$.

The fact that this method cannot always find real roots without using complex numbers (the *Casus Irreducibilis*) apparently caused difficulty in Cardano's era—that time before complex numbers. Again, à la Hadamard, the shortest route visits the complex numbers.

We wish again to emphasize that the possibly mysterious appearance of ω in the roots $a + b$, $a\omega + b\overline{\omega}$, $a\overline{\omega} + b\omega$ of $p(z)$ is actually just due to the fact that those roots are the images of the third roots of unity under the linear map $az + b\overline{z}$.

Higher dimensions

Informally, a regular tetrahedron can be rotated and dilated so that the first coordinates of its vertices match any four given numbers. Motivated by FIGURE 3, we generalize up one dimension (FIGURE 6) and state the result as a theorem (with proof left to the reader).

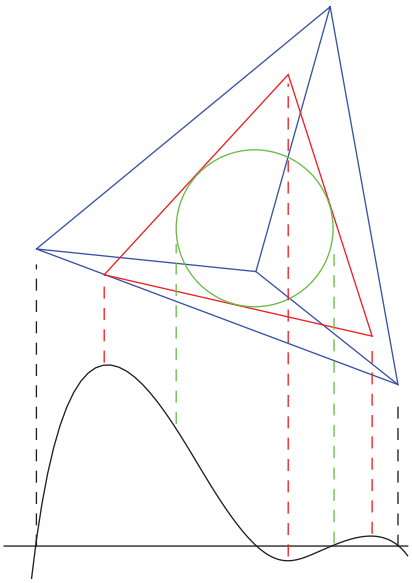


Figure 6 Tetrahedron, its inscribed sphere, and a triangle

THEOREM. *Given a quartic $p(x)$ with four real roots (at least two distinct), those roots are the first coordinate projections of a regular tetrahedron in $\mathbb{R}^3$. That tetrahedron has a unique inscribed sphere, which projects onto an interval whose endpoints are the two roots of $p''(x)$.*

FIGURE 6 shows the graph of a quartic polynomial $p(z)$ with a regular tetrahedron, whose vertices project onto the roots of $p(z)$; the inscribed sphere of that tetrahedron, whose extreme points project onto the roots of $p''(z)$; and an equilateral triangle around that sphere, whose vertices project onto the roots of $p'(z)$. Alas, FIGURE 6 does not indicate how the roots of $p'(z)$ relate to the roots of $p(z)$ or of $p''(z)$. If we could understand how *all* these roots are related geometrically, then we could approach the following deceptively simple open problem [**8**]. See [**2**] for a positive solution in lower dimensions.

CONJECTURE. *There does not exist a quartic polynomial p with four distinct rational roots such that p', p'', and p''' all have rational roots.*

Another surprising result involving FIGURE 6 is that the four points representing the vertices of the tetrahedron (as four complex numbers a, b, c, d in the complex plane on which FIGURE 6 is drawn) satisfy $(a^2 + b^2 + c^2 + d^2)/4 = ((a + b + c + d)/4)^2$ (the square of the average is the average of the squares). This, in fact, characterizes regular tetrahedra—proof left to the reader!

Another open problem about the relation between the roots of a polynomial p and the roots of its derivative p' is Sendov's conjecture, open since 1958 [**7**]. The degree 3 case follows from Marden's theorem and several higher-dimensional cases are known.

CONJECTURE. *Let p be a polynomial with roots $\{r_1, \ldots, r_n\}$ and let $\{s_1, \ldots, s_m\}$ be the roots of p'. Then*

$$\max_k \min_j |s_j - r_k| \leq \max_k |r_k|.$$

Geometrically, this says that if the roots of p are in the unit disk, then every root of p' is in a unit disk centered at some root of p.

REFERENCES

1. W. Dunham, *Journey Through Genius, The Great Theorems of Mathematics,* John Wiley, New York, 1990.
2. S. Gupta and W. Szymanski, Cubic polynomials with rational roots and critical points, *College Math. J.* **41** (2010) 365–369. http://dx.doi.org/10.4169/074683410X521964
3. P. Henrici, *Applied and Computational Complex Analysis, Vol. 1*, John Wiley, New York, 1986.
4. D. Kalman, An elementary proof of Marden's theorem, *Amer. Math. Monthly* **115** (2008) 330–338.
5. I. Kra and S. R. Simanca, On circulant matrices, *Notices of the AMS* **59** (2012) 368–377. http://dx.doi.org/10.1090/noti804
6. R. W. D. Nickalls, A new approach to solving the cubic: Cardano's solution revealed, *Math. Gazette* **77** no. 480 (Nov. 1993) 354–359. http://dx.doi.org/10.2307/3619777
7. Q. I. Rahman and G. Schmeisser, *Analytic Theory of Polynomials*, Oxford University Press, Oxford, 2002.
8. Quartic Rationally Derived Polynomials, *Open Problem Garden,* http://garden.irmacs.sfu.ca.
9. C. Taylor, *An Introduction to the Ancient and Modern Geometry of Conics,* Deighton Bell, Cambridge, 1881.
10. E. Weisstein, *Cubic formula*, from MathWorld, A Wolfram Web Resource, http://mathworld.wolfram.com/CubicFormula.html.

Summary Imagine a sphere with its equator inscribed in an equilateral triangle. This Saturn-like figure will help us understand from where Cardano's formula for finding the roots of a cubic polynomial $p(z)$ comes. It will also help us find a new proof of Marden's theorem, the surprising result that the roots of the derivative $p'(z)$ are the foci of the ellipse inscribed in and tangent to the midpoints of the triangle determined by the roots of the polynomial.

What Is Special about the Divisors of 12?

SUNIL K. CHEBOLU
Illinois State University
Normal, IL 61790
schebol@ilstu.edu

MICHAEL MAYERS
Illinois State University
Normal, IL 61790
mikemayers@gmail.com

This paper is a sequel to "What is special about the divisors of 24?" [**2**], in which the first author answered the following question, which evolved out of a classroom discussion. For what values of n does the multiplication table for $\mathbb{Z}_n$ have 1's only on the diagonal? In other words, when does $\mathbb{Z}_n$ have the property that whenever $ab = 1$, $a = b$? It was shown that only the divisors of 24 have this so-called diagonal property. In fact, that paper gives five proofs of this result. In its last section, the following variation of the above question was posed. *For what values of n does the multiplication table for $\mathbb{Z}_n[x]$ have 1's only on the diagonal?* We will solve this problem here. The problem is meaningful even though the multiplication table for $\mathbb{Z}_n[x]$ has infinite size. In fact, we can study this question over an arbitrary ring.

Let R be a commutative ring with multiplicative identity 1. The ring R is said to have the *diagonal property* if $a = b$ whenever $ab = 1$ in R. Before going further, we recall some standard definitions. Let a be an element of R. Then

- a is a *unit* if there exists b in R with $ab = 1$. (Then b is called the *inverse* of a.)
- a is an *involution* if it is its own inverse (or equivalently, a unit with $a^2 = 1$).

Math. Mag. **86** (2013) 143–146. doi:10.4169/math.mag.86.2.143.

- a is *nilpotent* if for some positive integer k, $a^k = 0$.

The diagonal property of R is equivalent to the following algebraic statement: Every unit in R is an involution. To see this, let R be a ring with the diagonal property and let a be a unit in R. Then, by definition of a unit, there is an element b in R such that $ab = 1$. Since R has the diagonal property, $a = b$. This means $a^2 = 1$, so a is an involution. For the other direction, suppose that every unit in R is an involution. If $ab = 1$, then a (being a unit) is also an involution. That is, $a^2 = 1$. Combining the last two equations, we get $ab = a^2$, and therefore $a = b$. This means that R has the diagonal property.

We will prove the following general result, which gives a surprising answer to the above question.

THEOREM. *For any positive integer m, the multiplication table for the polynomial ring* $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$ *has 1's only on the diagonal if and only if n is a divisor of* 12.

Proof. We will use basic facts from commutative algebra, which can be found in standard textbooks [**1, 3**].

Fix an arbitrary positive integer m, which will be the number of variables in our polynomial ring. We begin with a straightforward observation. Suppose n is a positive integer for which $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$ has the diagonal property. Since $\mathbb{Z}_n$ is a subring of $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$, it follows that $\mathbb{Z}_n$ also has the diagonal property. So n has to be a divisor of 24; see [**2**].

Now 8 and 24 are the only numbers that divide 24 but not 12. So our main theorem will follow once we prove the following statements.

(a) $\mathbb{Z}_8[x_1, x_2, \ldots, x_m]$ and $\mathbb{Z}_{24}[x_1, x_2, \ldots, x_m]$ do not have the diagonal property.

(b) $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$ has the diagonal property when n is a divisor of 24 that is not 8 or 24.

Proof of (a) In light of the above discussion, it is enough to find, in both rings, a unit that is not an involution. To this end, we will use the following well-known lemma.

LEMMA. *Let R be a commutative ring. If u is a unit and r is a nilpotent element in R, then* $u + r$ *is a unit.*

Proof. Let k be the unique integer such that $r^k \neq 0$ and $r^{k+1} = 0$. Then an easy check shows that the formal inverse of $u + r$ is given by

$$(1/u)(1 - r/u + (r/u)^2 + \cdots + (-1)^k(r/u)^k). \qquad \blacksquare$$

Returning to the proof of (a):

$\mathbb{Z}_8[x_1, x_2, \ldots, x_m]$: The element $2x_1$ is nilpotent in this ring and therefore, by the above lemma, $u = 1 + 2x_1$ is a unit. However,

$$u^2 - 1 = (1 + 2x_1)^2 - 1 = 4x_1 + 4x_1^2$$

is not a multiple of 8 in $\mathbb{Z}[x_1, x_2, \ldots, x_m]$, and hence not zero in $\mathbb{Z}_8[x_1, x_2, \ldots, x_m]$. Therefore u is not an involution.

$\mathbb{Z}_{24}[x_1, x_2, \ldots, x_m]$: The element $6x_1$ is nilpotent, and therefore $u = 1 + 6x_1$ is a unit. But

$$u^2 - 1 = (1 + 6x_1)^2 - 1 = 12x_1 + 36x_1^2$$

is not a multiple of 24 in $\mathbb{Z}[x_1, x_2, \ldots, x_m]$, and hence not zero in $\mathbb{Z}_{24}[x_1, x_2, \ldots, x_m]$. Therefore u is not an involution.

This completes the proof of part (a).

Proof of (b) Here we will use the following commutative algebra result, which gives a characterization of units in polynomial rings. Although this proposition is well-known, it is often stated only in the one-variable case.

PROPOSITION. *Let R be a commutative ring. A polynomial $f(x_1, x_2, \ldots, x_m)$ is a unit in $R[x_1, x_2, \ldots, x_m]$ if and only if the constant term of f is a unit in R and all other coefficients of f are nilpotent elements in R.*

Proof. We will use the fact [**1**] that an element in R is nilpotent if and only if it is in the intersection of all prime ideals of R.

The proof of the "if" direction follows easily from the above lemma and induction. As for the "only if" direction, we first prove it in the special case when R is an integral domain. To this end, suppose that R is an integral domain and let f (as above) be a unit in $R[x_1, x_2, \ldots, x_m]$. Then there exists a polynomial g such that $fg = 1$. Since R is an integral domain, we have $\deg(fg) = \deg(f) + \deg(g)$. So we get

$$\deg(f) + \deg(g) = 0.$$

This means that f and g are constant terms. Since $fg = 1$, it follows that f is a unit in R.

For the general case, let f be a unit in $R[x_1, x_2, \ldots, x_m]$ and consider the ring homomorphism

$$R[x_1, x_2, \ldots, x_m] \longrightarrow R,$$

which reduces modulo the ideal $(x_1, x_2, \ldots, x_m)$. The image of f under this homomorphism is the constant term of f. Since every ring homomorphism sends units to units, we conclude that the constant term of f is a unit. To see that the other coefficients of f are nilpotent elements in R, we let p be any prime ideal in R, and consider the ring homomorphism

$$R[x_1, x_2, \ldots, x_m] \longrightarrow R/p[x_1, x_2, \ldots, x_m],$$

which reduces the coefficients of a polynomial modulo p. Since R/p is an integral domain, by the special case proved above, we conclude that the image of f in $R/p[x_1, x_2, \ldots, x_m]$ is a constant, i.e., the coefficients of all higher-degree terms are zero. This means that they all belong to p. Since the choice of prime ideal p was arbitrary, it follows that all the coefficients of f, other than the constant term, belong to the intersection of all prime ideals. By the aforementioned fact, the latter is precisely the set of nilpotent elements of R, and therefore the coefficients in question are nilpotent. ∎

Returning to the proof of (b):

We will show that whenever n is a divisor of 24 other than 8 or 24, every unit u in $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$ is an involution, i.e., $u^2 = 1$. The values of n to be considered are 2, 3, 4, 6, and 12.

The rings $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_6$ are reduced. That is, they do not have any nonzero nilpotent elements. Therefore, by the above proposition, the units in $\mathbb{Z}_2[x_1, x_2, \ldots, x_m]$, $\mathbb{Z}_3[x_1, x_2, \ldots, x_m]$, and $\mathbb{Z}_6[x_1, x_2, \ldots, x_m]$ are exactly those in $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_6$, respectively. Since the rings $\mathbb{Z}_2$, $\mathbb{Z}_3$, and $\mathbb{Z}_6$ have the diagonal property, so do the corresponding polynomial rings. This leaves us with 4 and 12.

$\mathbb{Z}_4[x_1, x_2, \ldots, x_m]$: 2 is the only nonzero nilpotent and 1 and -1 are the only units in $\mathbb{Z}_4$. Therefore, every unit u in $\mathbb{Z}_4[x_1, x_2, \ldots, x_m]$ is of the form $2h - 1$ or $2h + 1$, where h is an arbitrary polynomial. Then $u^2 = 1 \mod 4$ in $\mathbb{Z}[x_1, x_2, \ldots, x_m]$, and therefore u is an involution in $\mathbb{Z}_4[x_1, x_2, \ldots, x_m]$.

$\mathbb{Z}_{12}[x_1, x_2, \ldots, x_m]$: 6 is the only nonzero nilpotent in $\mathbb{Z}_{12}$. Therefore, every unit u in $\mathbb{Z}_{12}[x_1, x_2, \ldots, x_m]$ is of the form $u = 6h + r$, where h is an arbitrary polynomial and r is a unit in $\mathbb{Z}_{12}$. Now, $u^2 = 36h^2 + 12rh + r^2$. The latter is equal to r^2 in $\mathbb{Z}_{12}[x_1, x_2, \ldots, x_m]$. Since $\mathbb{Z}_{12}$ has the diagonal property, we have $r^2 = 1 \mod 12$. Therefore u is an involution in $\mathbb{Z}_{12}[x_1, x_2, \ldots, x_m]$, as desired. ■

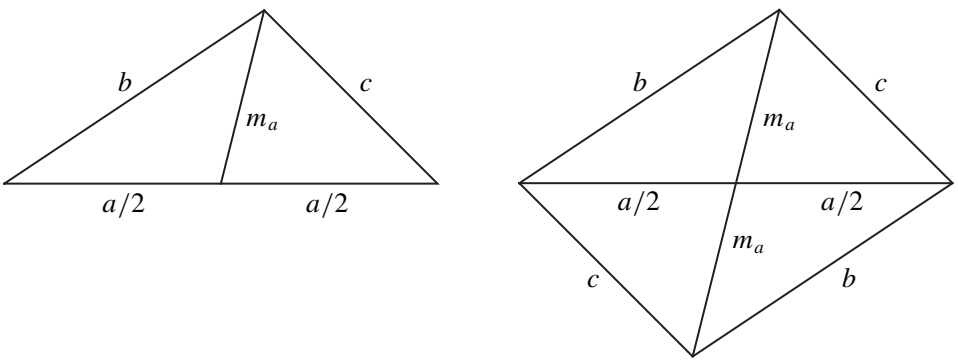
REFERENCES

1. M. F. Atiyah and I. G. MacDonald, *Introduction to Commutative Algebra,* Addison-Wesley, Reading, MA, 1969.
2. Sunil K. Chebolu, What is special about the divisors of 24?, *Math. Mag. 85* (2012) 366–372.
3. David S. Dummit and Richard M. Foote, *Abstract Algebra,* 3rd ed., John Wiley, New York, 2004.

Summary What is an interesting number theoretic characterization of the divisors of 12 among all positive integers? This paper will provide one answer in terms of modular multiplication tables. We will show that the multiplication table for the ring $\mathbb{Z}_n[x_1, x_2, \ldots, x_m]$ has 1's only on the diagonal if and only if n is a divisor of 12.

Proof Without Words: The Length of a Triangle Median via the Parallelogram Law

C. PETER LAWES
Northern Illinois University
DeKalb, IL 60115
cplawes@comcast.net



$$2b^2 + 2c^2 = a^2 + (2m_a)^2$$

$$\therefore \qquad m_a = \frac{1}{2}\sqrt{2\left(b^2 + c^2\right) - a^2}$$

Math. Mag. **86** (2013) 146. doi:10.4169/math.mag.86.2.146.

PROBLEMS

BERNARDO M. ÁBREGO, *Editor*
California State University, Northridge

Assistant Editors: SILVIA FERNÁNDEZ-MERCHANT, California State University, Northridge; JOSÉ A. GÓMEZ, Facultad de Ciencias, UNAM, México; EUGEN J. IONASCU, Columbus State University; ROGELIO VALDEZ, Facultad de Ciencias, UAEM, México; WILLIAM WATKINS, California State University, Northridge

PROPOSALS

To be considered for publication, solutions should be received by September 1, 2013.

1916. *Proposed by H. A. ShahAli, Tehran, Iran.*

Let M and n be positive integers. For every integer d, let r_d be the remainder obtained when n is divided by d. Let $R_n = \{r_d : 1 \le d \le \lfloor n/2 \rfloor\}$. Prove that $|R_n| > M$ holds for all sufficiently large positive integers n.

1917. *Proposed by Ovidiu Furdui, Technical University of Cluj-Napoca, Cluj, Romania.*

Let $\alpha > 0$ and let a and b be real numbers such that $b > a$. Find the value of

$$\lim_{n\to\infty} \int_a^b \sqrt[n]{(x-a)^n + (b-x)^n + \alpha((x-a)(b-x))^{n/2}}\, dx.$$

1918. *Proposed by Kent Holing, Trondheim, Norway.*

A triangle is given with side lengths a, b, c, and with inradius and circumradius r and R, respectively. Let $s = \frac{1}{2}(a+b+c)$ denote the semiperimeter of the triangle.

(a) Solve the cubic $x^3 - (r+4R)x^2 + s^2x - rs^2 = 0$ exactly by starting with the monic cubic having roots $s-a$, $s-b$, and $s-c$.

(b) Using (a), give a geometrical interpretation of the roots of the given cubic.

1919. *Proposed by Abdurrahim Yilmaz, Middle East Technical University, Ankara, Turkey.*

Let $n \ge 3$ be a positive integer. A circle of radius 1 centered at the origin is inscribed by a circular arrangement of n congruent circles, where every two consecutive circles in the arrangement are tangent. This arrangement is in turn inscribed by a second cir-

Math. Mag. **86** (2013) 147–154. doi:10.4169/math.mag.86.2.147.

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any 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.

Solutions should be written in a style appropriate for this MAGAZINE.

Solutions and new proposals should be mailed to Bernardo M. Ábrego, Problems Editor, Department of Mathematics, California State University, Northridge, 18111 Nordhoff St, Northridge, CA 91330-8313, or mailed electronically (ideally as a LaTeX or pdf file) to mathmagproblems@csun.edu. All communications, written or electronic, should include **on each page** the reader's name, full address, and an e-mail address and/or FAX number.

cular arrangement of n congruent circles in two different ways, as shown in the figures below. In Figure 1, the line through the centers of tangent circles in different arrangements goes through the origin. In Figure 2, each circle in the second arrangement is tangent to two consecutive circles in the first arrangement. Suppose that this process continues indefinitely as shown in the figures. For each of the two cases, find the area enclosed by the union of all circles when $n \to \infty$.

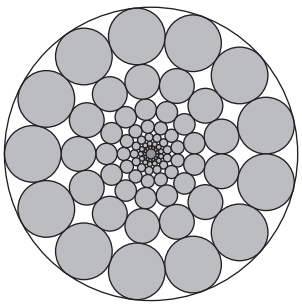


Figure 1

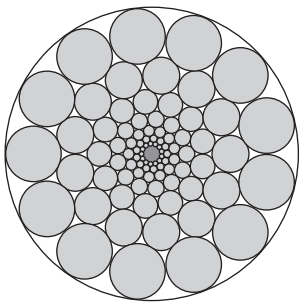


Figure 2

1920. *Proposed by Iliya Bluskov, University of Northern British Columbia, Prince George, Canada.*

For every positive integer n, prove that

$$\sum_{m=0}^{n}\sum_{r=0}^{n-m}\frac{(-2)^r}{m!r!}\binom{2n-m-r}{n}=\sum_{r=0}^{n}\frac{(-1)^r}{r!}\binom{2n-r}{n}.$$

Quickies

Answers to the Quickies are on page 153.

Q1029. *Proposed by Tom Moore, Bridgewater State University, Bridgewater, MA.*

For what fields F is it true that $-a = a^{-1}$ for all its nonzero elements?

Q1030. *Proposed by Alina Sîntămărian, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.*

Let $p \geq 2$ be an integer and let $(x_n)_{n\in\mathbb{N}}$ be the sequence defined by

$$x_n = \prod_{j=0}^{n-1}\frac{n+jp+1}{n+jp}.$$

Assuming that $\lim_{n\to\infty} x_n$ exists, find its value.

Solutions

The homothetic quadrilateral of centroids — April 2012

1891. *Proposed by Raúl A. Simón, Santiago, Chile.*

Let $ABCD$ be a quadrilateral in the plane. Let A', B', C', and D' be the centroids of the triangles BCD, ACD, ABD, and ABC, respectively. Prove that quadrilaterals $ABCD$ and $A'B'C'D'$ are similar with corresponding sides in the ratio $3:1$.

I. *Solution by Dmitry Fleischman, Santa Monica, CA.*
Let M be the midpoint of $\overline{CD}$. The point A' divides the median $\overline{BM}$ in the ratio 2 : 1 and point B' divides the median $\overline{AM}$ in the ratio 2 : 1. It follows that $\triangle ABM$ is similar to $\triangle B'A'M$ in the ratio 3 : 1. Thus, $\overline{A'B'}$ is parallel to $\overline{AB}$ and $A'B' = \frac{1}{3}AB$. Similarly, each of the other sides of $A'B'C'D'$ is parallel to, and measures a third of the length of, the corresponding side of $ABCD$.

II. *Solution by Mowaffaq Hajja and Mostafa Hayajneh, Yarmouk University, Irbid, Jordan.*
Place the quadrilateral $ABCD$ in the plane so that its centroid lies at the origin O. In other words, assume that $A + B + C + D = O$. Then the centroid A' of triangle BCD is given by $A' = \frac{1}{3}(B + C + D) = -\frac{1}{3}A$. Similarly, $B' = -\frac{1}{3}B$, $C' = -\frac{1}{3}C$, and $D' = -\frac{1}{3}D$. Thus $A'B'C'D'$ is homotethic to $ABCD$ with homotethy center O and homotethy ratio $-\frac{1}{3}$.

Note that the same proof applies to any number $n \geq 2$ of points $A_1, \ldots, A_n$ lying in $\mathbb{R}^d$ for any $d \geq 1$. That is, if the center of mass of $A_1 \ldots A_n$ is the origin O, and A'_j denotes the center of mass of $A_1 \ldots A_{j-1}A_{j+1} \ldots A_n$, then $A'_j = -(1/(n-1))A_j$. It follows that $A'_1 \ldots A'_n$ is homotethic to $A_1 \ldots A_n$, with homotethy center O and homotethy ratio $-1/(n-1)$.

Editor's Note. M. J. Englefield notes that this problem and its solution has appeared in [John Alison, Statical proofs of some geometrical theorems, *Proc. Edinburgh Math. Soc.* IV (1885) 58–60.]

Also solved by Armstrong Problem Solvers, Herb Bailey, Michel Bataille (France), Jany C. Binz (Switzerland), Elton Bojaxhiu (Albania) and Enkel Hysnelaj (Australia), Bruce S. Burdick, Robert Calcaterra, Julio Castiñeira Merino (Spain), Tim Cross (United Kingdom), Chip Curtis, Prithwijit De (India), Marian Dincă (Romania), M. J. Englefield (Australia), Thomas Gettys, Michael Goldenberg and Mark Kaplan, Eugene A. Herman, Omran Kouba (Syria), Victor Y. Kutsenok, Elias Lampakis (Greece), Robert W. Langer, Deokjae Lee (Korea), Graham Lord, Missouri State University Problem Solving Group, José Heber Nieto (Venezuela), Peter Nüesch (Switzerland), Ángel Plaza (Spain) and Javier Sánchez-Reyes (Spain), Juyeop Song (Korea), H. T. Tang, Traian Viteam (Uruguay), Michael Vowe (Switzerland), Michael Woltermann, John B. Zacharias, Tom Zerger, and the proposer. There was one incorrect submission.

Products turned into Riemann sums **April 2012**

1892. *Proposed by José Luis Díaz-Barrero, Applied Mathematics III, Polytechnical University of Catalonia, Barcelona, Spain.*

Compute the limit

$$\lim_{n\to\infty} \frac{1}{n^n} \prod_{k=1}^{n} \left(\frac{n\sqrt{n} + (n+1)\sqrt{k}}{\sqrt{n} + \sqrt{k}} \right).$$

Solution by Traian Viteam, Montevideo, Uruguay.
Let us show more generally that, for a continuous function $f : [0, 1] \to [0, \infty)$ (a Riemann integrable sum would suffice), we have

$$P_n(f) = \prod_{k=1}^{n} \left(1 + \frac{1}{n} f\left(\frac{k}{n}\right) \right) \underset{n\to\infty}{\longrightarrow} e^{\int_0^1 f(x)dx}. \tag{1}$$

Then, the required limit equals

$$\lim_{n\to\infty} \prod_{k=1}^{n} \left(\frac{\sqrt{n} + (1+1/n)\sqrt{k}}{\sqrt{n} + \sqrt{k}} \right) = \lim_{n\to\infty} \prod_{k=1}^{n} \left(1 + \frac{1}{n} \frac{\sqrt{k/n}}{1 + \sqrt{k/n}} \right) = \frac{4}{e},$$

by letting $f(x) = f_0(x) = \sqrt{x}/(1+\sqrt{x})$ in (1), and observing that

$$\int_0^1 f_0(x)\,dx = \int_0^1 \frac{2u^2}{1+u}\,du = 2\int_0^1 \left(u - 1 + \frac{1}{u+1}\right)du = 2\ln 2 - 1.$$

To show (1), we first observe that if $x \geq 0$, then

$$x - \ln(1+x) = \int_1^{1+x} \frac{t-1}{t}\,dt,$$

and

$$0 \leq \int_1^{1+x} \frac{t-1}{t}\,dt \leq \int_1^{1+x} (t-1)\,dt = \frac{x^2}{2}.$$

This implies that for every $n \geq 1$,

$$0 \leq \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right) - \ln P_n(f) \leq \frac{1}{2n} \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right)^2.$$

The definition of Riemann integral ensures that $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right) = \int_0^1 f(x)\,dx$ and $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right)^2 = \int_0^1 f(x)^2\,dx$. Letting $n \to \infty$ gives

$$0 \leq \int_0^1 f(x)\,dx - \lim_{n\to\infty} \ln P_n(f) \leq \lim_{n\to\infty} \frac{1}{2n} \int_0^1 f(x)^2\,dx = 0.$$

Thus $\lim_{n\to\infty} \ln P_n(f) = \int_0^1 f(x)\,dx$ and (1) is proved.

Editor's Note. Using similar arguments, Edward Omey showed that if f' exists and it is integrable then the following asymptotic formula is true

$$P_n(f) = e^{\int_0^1 f(x)dx}\left(1 + \frac{f(1)-f(0)}{2n} - \frac{1}{2n}\int_0^1 f(x)^2 dx + \Theta\left(\frac{1}{n^2}\right)\right),$$

which in this case gives

$$\lim_{n\to\infty} n\left(P_n(f_0) - \frac{4}{e}\right) = \frac{9 - 12\ln 2}{e}.$$

Also solved by Robert A. Agnew, Michel Bataille (France), Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Germany), Robert Calcaterra, Chip Curtis, Eugene S. Eyeson, John N. Fitch, William R. Green, Eugene A. Herman, Stephen Kaczkowski, Omran Kouba (Syria), Elias Lampakis (Greece), Reiner Martin (Germany), Peter McPolin (Northern Ireland), Northwestern University Math Problem Solving Group, Edward Omey (Belgium), Paolo Perfetti (Italy), Tomas Persson (Sweden) and Mikael P. Sundqvist (Sweden), Nicholas C. Singer, Dave Trautman, Tiberiu Trif (Romania), Haohao Wang and Jerzy Woydyło, José Heber Nieto (Venezuela), John Zacharias, and the proposer. There was one incomplete solution.

On the product of pairwise differences **April 2012**

1893. *Proposed by Jerrold W. Grossman and László Lipták, Oakland University, Rochester, MI, and Mike Shaughnessy, Portland State University, Portland, OR.*

Let n be an integer greater than 1, and let

$$f(a_1, a_2, \ldots, a_n) = \prod_{1\leq i<j\leq n} (a_i - a_j).$$

What is the greatest common divisor of $f(a_1, a_2, \ldots, a_n)$ over all choices of distinct integers $a_1, a_2, \ldots, a_n$?

Solution by Michel Bataille, Rouen, France.

We show that the required greatest common divisor is $D = \prod_{k=1}^{n-1} k!$. We observe that $D = f(n, n-1, \ldots, 2, 1)$, so it is sufficient to prove that D divides $f(a_1, a_2, \ldots, a_n)$ for all n-tuples of integers $(a_1, a_2, \ldots, a_n)$. Notice that, up to sign, $f(a_1, a_2, \ldots, a_n)$ is the determinant of the Vandermonde matrix

$$V = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{pmatrix}.$$

Let $P_1, P_2, \ldots, P_{n-1}$ be monic polynomials with integral coefficients and degrees $1, 2, \ldots, n-1$, respectively. Note that the matrix

$$V' = \begin{pmatrix} 1 & P_1(a_1) & P_2(a_1) & \cdots & P_{n-1}(a_1) \\ 1 & P_1(a_2) & P_2(a_2) & \cdots & P_{n-1}(a_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & P_1(a_n) & P_2(a_n) & \cdots & P_{n-1}(a_n) \end{pmatrix}$$

can be obtained by adding to each column of V, except the first, a suitable linear combination of the preceding ones (starting with the last column). Because all of these elementary column operations preserve the determinant, it follows that $\det V' = \det V$.

Doing this with the monic polynomials

$$\begin{aligned} P_1(x) &= x, \\ P_2(x) &= x(x+1), \ldots, \\ P_{n-1}(x) &= x(x+1)\cdots(x+n-2), \end{aligned}$$

each entry in column k $(2 \le k \le n)$ is a product of $k-1$ consecutive positive integers and, as such, is divisible by $(k-1)!$ (since $a(a+1)\cdots(a+k-2) = (k-1)!\binom{a+k-2}{k-1}$). Extracting each of these factors from the determinant gives

$$\det V = \det V' = (1!\,2!\,3!\cdots(n-1)!)\det V_1,$$

where V_1 is a matrix of integers. Because $\det V_1$ is an integer, the result follows.

Editor's Note. The result of this problem is not new. Some readers mentioned [B. Sury, An integral polynomial, this *Magazine* **68** (1995) 134–135] and [R. Chapman, A polynomial taking integer values, this *Magazine* **69** (1996) 121] as relevant references where this problem is solved. Traian Viteam notes that problem E2637 in *Amer. Math. Monthly* asks to prove that $\prod_{i,j}(a_i - a_j)/(i-j)$ is an integer. The solution to E2637 in [*Amer. Math. Monthly* **85** (1978) 386–387] contains more references to other solutions, some from as early as 1881. That solution proves that, up to sign, $\prod_{i,j}(a_i - a_j)/(i-j)$ is the determinant of the $n \times n$ matrix with entries $\binom{a_j}{i-1}$. As a consequence of this, we can obtain the following factorization:

$$\prod_{1 \le i < j \le n} (a_i - a_j) = \prod_{1 \le i < j \le n} (i-j) \cdot \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \binom{a_i}{\sigma(i) - 1}.$$

Also solved by Elton Bojaxhiu (Albania) and Enkel Hysnelaj (Australia), Robert Calcaterra, Con Amore Problem Group (Denmark), Dmitry Fleischman, Tae Gyun Kim (Korea), Elias Lampakis (Greece), José Heber Nieto (Venezuela), Traian Viteam (Uruguay), and the proposers. There were two incorrect submissions.

The isolated points of a set **April 2012**

1894. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let $(X, \mathfrak{T})$ be a topological space, let S be any subset of X, and let $\text{iso}(S)$ be the set of isolated points of S. (A point $x \in S$ is an isolated point of S if there exists an open set O such that $O \cap S = \{x\}$.)

(a) If $(X, \mathfrak{T})$ is second countable, prove that $\text{iso}(S)$ is countable. (A topological space $(X, \mathfrak{T})$ is second countable if it has a countable open base.)

(b) If S is a closed set, prove that $S \setminus \text{iso}(S)$ is a closed set.

(c) Prove that the conclusion in (a) does not hold if $(X, \mathfrak{T})$ is only assumed to be separable. (A topological space $(X, \mathfrak{T})$ is separable if it has a countable dense subset.)

(d) Prove that the conclusion in (b) does not hold if S is not assumed to be closed.

Solution by Eugene Herman, Grinnell College, Grinnell, IA.

(a) Let B be a countable open base for $(X, \mathfrak{T})$. For each $x \in \text{iso}(S)$, choose O_x open such that $O_x \cap S = \{x\}$. Then choose $U_x \in B$ such that $x \in U_x \subseteq O_x$, and so $U_x \cap S = \{x\}$. Since the map $x \mapsto U_x$ from $\text{iso}(S)$ to B is injective, $\text{iso}(S)$ is countable.

(b) For each $x \in \text{iso}(S)$, choose O_x open such that $O_x \cap S = \{x\}$, and let $O = \cup_{x \in \text{iso}(S)} O_x$. Hence O is open and $O \cap S = \text{iso}(S)$. Therefore $(S \setminus \text{iso}(S))^c = \text{iso}(S) \cup S^c = (O \cap S) \cup S^c = O \cup S^c$. Since $O \cup S^c$ is open, $S \setminus \text{iso}(S)$ is closed.

(c) Let X be any uncountable set and $x_0 \in X$. Let $O \subseteq X$ be open if and only if O is empty or $x_0 \in O$. This defines a separable topology on X, since $\{x_0\}$ is dense in X. Let $S = X \setminus \{x_0\}$. Then S is uncountable, and each $x \in S$ is isolated since $\{x, x_0\}$ is open.

(d) If S is any set that is not closed and has no isolated points, then $S \setminus \text{iso}(S)$ is not closed. For example, let $S = [0, 1)$ on the real line with the usual topology.

Also solved by John Atkins, Michel Bataille (France), Douglas K. Brown, Paul Budney, Bruce S. Burdick, Robert Calcaterra, Michelle Daher (Lebanon), Laura Iosip (United Kingdom), Elias Lampakis (Greece), Kathleen E. Lewis (Republic of the Gambia), Richard P. Millspaugh, Charlie Mumma, José Heber Nieto (Venezuela), Paolo Perfetti (Italy), Ángel Plaza (Spain) and Kishin Sadarangani (Spain), Tiberiu Trif (Romania), and the proposer.

The probability density function of the maximum angle **April 2012**

1895. *Proposed by Steven Finch, Harvard University, Cambridge, MA*

Let ℓ denote a planar line with slope $\tan(\theta)$ and x-intercept ξ, where θ and ξ are independent random variables with uniform distributions over the intervals $[\pi/4, 3\pi/4]$ and $[-1, 1]$, respectively. Let ℓ_1, ℓ_2, and ℓ_3 be independent copies of ℓ. These three lines determine a compact triangle Δ almost surely. Find the probability density function for the maximum angle α in Δ. Find the first and second moments of α as well.

Solution by Elton Bojaxhiu, Kriftel, Germany, and Enkel Hysnelaj, University of Technology, Sydney, Australia.

Let θ_1, θ_2, and θ_3 be the slope angles of ℓ_1, ℓ_2, and ℓ_3, respectively. They are independent and they have the same distributions, so $P(\theta_1 < \theta_2 < \theta_3) = 1/3! = 1/6$. If $\theta_1 < \theta_2 < \theta_3$, then the triangle formed by the lines ℓ_1, ℓ_2, and ℓ_3 will have angles $\theta_2 - \theta_1$, $\theta_3 - \theta_2$, and $\pi - \theta_3 + \theta_1$. Since the third angle satisfies that $\pi/2 = \pi - 3\pi/4 + \pi/4 \leq$

$\pi - \theta_3 + \theta_1 \leq \pi$, then it will be the maximal angle $\alpha = \pi - \theta_3 + \theta_1$. Let us first compute the distribution function of α. For $\pi/2 \leq \theta \leq \pi$,

$$\begin{aligned}
P(\alpha < \theta) &= P(\pi - \theta_3 + \theta_1 < \theta \mid \theta_1 < \theta_2 < \theta_3) \\
&= \frac{P(\theta_3 > \theta_1 + \pi - \theta, \theta_1 < \theta_2, \theta_2 < \theta_3)}{P(\theta_1 < \theta_2 < \theta_3)} \\
&= 6P(\theta_3 > \max(\theta_2, \theta_1 + \pi - \theta), \theta_2 > \theta_1) \\
&= \frac{6}{\left(\frac{3\pi}{4} - \frac{\pi}{4}\right)^3} \int_{\frac{\pi}{4}}^{\theta - \frac{\pi}{4}} \left[\int_{\theta_1}^{\theta_1 + \pi - \theta} d\theta_2 \int_{\theta_1 + \pi - \theta}^{\frac{3\pi}{4}} d\theta_3 \right. \\
&\qquad \left. + \int_{\theta_1 + \pi - \theta}^{\frac{3\pi}{4}} \left(\int_{\theta_2}^{\frac{3\pi}{4}} d\theta_3 \right) d\theta_2 \right] d\theta_1 \\
&= (5\pi - 4\theta)(2\theta - \pi)^2/\pi^3.
\end{aligned}$$

Since $\pi/2 \leq \alpha \leq \pi$, then

$$F_\alpha(\theta) = P(\alpha < \theta) = \begin{cases} 0 & \text{if } \theta < \pi/2, \\ (5\pi - 4\theta)(2\theta - \pi)^2/\pi^3 & \text{if } \pi/2 \leq \theta \leq \pi, \text{ and} \\ 1 & \text{if } \theta > \pi. \end{cases}$$

The function F_α is continuous on $\mathbb{R}$ and its derivative is continuous on $\mathbb{R}$, except perhaps for the points $\pi/2$ and π. Thus, the density of the random variable α exists and is achieved by building the derivative F'_α,

$$p_\alpha(\theta) = \begin{cases} 24(2\theta - \pi)(\pi - \theta)/\pi^3 & \text{if } \pi/2 < \theta < \pi, \text{ and} \\ 0 & \text{else.} \end{cases}$$

The first moment of α is

$$\int_{-\infty}^{\infty} \theta p_\alpha(\theta) d\theta = \frac{24}{\pi^3} \int_{\frac{\pi}{2}}^{\pi} \theta(2\theta - \pi)(\pi - \theta) d\theta = \frac{3}{4}\pi.$$

The second moment of α is

$$\int_{-\infty}^{\infty} \theta^2 p_\alpha(\theta) d\theta = \frac{24}{\pi^3} \int_{\frac{\pi}{2}}^{\pi} \theta^2(2\theta - \pi)(\pi - \theta) d\theta = \frac{23}{40}\pi^2.$$

Also solved by Robert Calcaterra, Peter McPolin (Northern Ireland), and the proposer.

Answers

Solutions to the Quickies from page 148.

A1029. Suppose that $-a = a^{-1}$ for all nonzero elements of F. Consider the unity of F, say $1_F \neq 0_F$. Then $-1_F = {1_F}^{-1} = 1_F$. Thus $2 \cdot 1_F = 1_F + 1_F = 1_F + (-1_F) = 0_F$, and so F has characteristic 2. Thus $-a = a$ for all nonzero a . Also, $-a = a^{-1}$ yields $-a^2 = 1_F$. Thus $a^2 = -1_F = 1_F$. Therefore, $(a + 1_F)^2 = a^2 + 2a + 1_F = 1_F + 0_F + 1_F = 2 \cdot 1_F = 0_F$. But F has no nonzero zero-divisors, so $a + 1_F = 0_F$ and $a = 1_F$ follows. Thus $F = \{0_F, 1_F\}$ and F is isomorphic to $\mathbb{F}_2$.

A1030. For $1 \le m \le p$, let $(x_n^{(m)})_{n\in\mathbb{N}}$ be the sequence defined by

$$x_n^{(m)} = \prod_{j=0}^{n-1} \frac{n + jp + m}{n + jp + m - 1}.$$

Note that $x_n^{(1)} = x_n$, and

$$x_{n+m}^{(1)} = x_n^{(m+1)} \prod_{k=1}^{m} \frac{n + (n + k - 1)p + m + 1}{n + (n + k - 1)p + m}.$$

Thus,

$$x_n^{(1)} x_{n+1}^{(1)} \cdots x_{n+p-1}^{(1)} = x_n^{(1)} x_n^{(2)} \cdots x_n^{(p)} \prod_{m=1}^{p-1} \prod_{k=1}^{m} \frac{n + (n + k - 1)p + m + 1}{n + (n + k - 1)p + m}.$$

However, the product $x_n^{(1)} x_n^{(2)} \cdots x_n^{(p)}$ telescopes as

$$x_n^{(1)} x_n^{(2)} \cdots x_n^{(p)} = \prod_{j=0}^{n-1} \prod_{m=1}^{p} \frac{n + jp + m}{n + jp + m - 1} = \prod_{j=0}^{n-1} \frac{n + (j + 1)p}{n + jp} = p + 1.$$

Therefore,

$$x_n^{(1)} x_{n+1}^{(1)} \cdots x_{n+p-1}^{(1)} = (p + 1) \prod_{m=1}^{p-1} \prod_{k=1}^{m} \frac{n + (n + k - 1)p + m + 1}{n + (n + k - 1)p + m}.$$

It follows that $(\lim_{n\to\infty} x_n)^p = p + 1$, and so $\lim_{n\to\infty} x_n = \sqrt[p]{p + 1}$.

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.

Wolchover, Natalie, In computers we trust? As math grows ever more complex, will computers reign?, Simons Science News (22 February 22, 2013), https://simonsfoundation.org/features/science-news/in-computers-we-trust/ .

The Simons Foundation, founded by a mathematician from SUNY Stony Brook who made a fortune as a hedge-fund manager, is dedicated to advancing research in mathematics and science. The Foundation sponsors lectures on recent developments, as well as feature articles (lately, every two or three weeks, not all on mathematics). The Simons site includes text of the articles, together with videos of the lectures and of interviews with "giants" (e.g., Lovász, Morawetz, Nash, Manin, Sally, Atiyah). Other recent articles have treated a slight improvement on Christofides' algorithm for the traveling sales problem, and "universality" (systems—such as melt ponds, the Internet, and buses in Cuernavaca—behaving like random matrices). The article cited above features Doron Zeilberger (Rutgers) and his co-author Shalosh B. Ekhad (a computer) and celebrates the ubiquity of "computerized math": experimental mathematics and computer verification of mathematical proofs.

Langville, Amy N., and Carl D. Meyer, *Who's #1? The Science of Rating and Ranking*, Princeton Univ. Press, 2012; xv + 247 pp, $29.95. ISBN 978-0-691-15422-0.

As the Olympics and the impetus for a playoff series for college football attest, there is a great drive to know who is #1, in whatever. (This drive no doubt diminishes afterward for those who aren't.) "In America, especially, we are evaluation-obsessed, which thereby makes us ranking-obsessed." Authors Langville and Meyer also wrote *Google's PageRank and Beyond: The Science of Search Engine Rankings* (2006), so here they continue their "obsession." The book presumes linear algebra, though the authors optimistically suggest that others read it anyway (surprisingly, this is a trade book, not a textbook). Included are well-known methods by Massey, Colley, Keener, and Elo; others by Markov and Sinkhorn; and still others that take into account point spreads, ties, weighting, and rank aggregation. Not included are methods based on Saaty's analytic hierarchy process (with use of a geometric mean, it is equivalent to the Massey method), logistic regression, Markov chains, or Monte Carlo simulations. But which ranking method is #1?

Van Brummelen, Glen, *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*, Princeton Univ. Press, 2013; xvi + 192 pp + 11 color plates, $35. ISBN 978-0-691-14892-2.

Did I study spherical trigonometry in high school? Maybe. Does anyone study it today? No. But it has appeal, as this elegant book attests. However, "This is not a scholarly work in the history of mathematics. It does not contain footnotes, does not profess to tell the whole story.... This is simply an appreciation of a beautiful lost subject...." There are a dozen or so exercises at the end of each chapter, and the author conveys both the practicality and the charm of the subject.

Math. Mag. **86** (2013) 155–156. doi:10.4169/math.mag.86.2.155.

Wu, Hung-Hsi, Assessment for the Common Core Mathematics Standards, *Journal of Mathematics Education at Teachers College* 3 (Spring-Summer 2012) 6–18.

The Common Core State Standards for Mathematics (CCSSM) have been adopted by almost all U.S. states. Author Wu (UC–Berkeley) concentrates on suggesting strategies to avoid "pitfalls" in assessing curricula that implement those standards. He criticizes continuing use of ill-posed problems about patterns (10, 13, 16, 19; what is the next number?), order of operations ($3 + 5 \times 7 = ?$), and about concepts undefined in school textbooks (rate, scale drawing). Wu laments the dominant "bipolar approach to mathematics instruction," in which pure mathematics is taught by rote (so students don't learn to reason with its concepts) while applied mathematics is taught as about solving problems: "how to use the tools but not why the tools are true." [But isn't that exactly the kind of practical training in "mathematics skills" that our pragmatic culture demands from "education"?] Current "high-stakes tests" drive this practice by assessing not knowledge of mathematical theorems but only the ability to apply them to solve problems. Wu realizes the need for quick and easy grading, as well as the consequent drawback, that such tests "are fundamentally too crude to measure different levels of excellence in mathematics achievement." He means that they cannot measure students' capability for the "sustained sequential thinking" that is necessary for a mathematical argument (and for much other reasoning). So, what to do? Wu recommends that such tests concentrate on basic competence rather than assessing excellence, so that they become "driver's license" tests in math: "pass it and forget it," with no pressure on teachers to have students get higher and higher scores. Unfortunately, "pass it and forget it" in another sense describes only too well what in fact happens with "mathematics skills" learned in isolation from mathematical thinking. [Thanks to Henry O. Pollak.]

Stewart, Ian, *The Mathematics of Life*, Basic Books, 2011; viii + 358 pp, $27.99, $15.99 (P), $15.99 (eBook). ISBN 978-0-465-02238-0, 978-0-465-03240-2, 978-0-465-02440-7.

"Biology will be the great mathematical frontier of the twenty-first century." Stewart identifies six revolutions in biology: the microscope, systematic taxonomy, the theory of evolution, the gene, DNA, and now… biomathematics. The mathematics is handled with a light touch (no equations, few graphs); the book is intended to inspire a popular audience, with a wide variety of mathematical aspects (the probability in genetics, patterns of stripes on animals, DNA knots, viruses as stellated polyhedra, and much more). Fortunately, the alliteration in the titles of chapters is (almost) exhausted after "Long List of Life" and "Florally Finding Fibonacci."

Slingerland, Rudy, and Lee Kump, *Mathematical Modeling of Earth's Dynamical Systems: A Primer*, Princeton University Press, 2011; xii + 231 pp, $45 (P). ISBN 978-0-691-14514-3.

I was interested in this book because I occasionally teach a course in environmental modeling that has no prerequisites. This book, though, is for advanced undergraduate and graduate students in the earth sciences who are "familiar with the principles of physics, chemistry, and geology" and have had a year of calculus; that last requirement, likely fulfilled several years earlier in the student's career, will not suffice, since the authors admit that the material also requires knowledge of ordinary and partial differential equations. The book cites actual situations and problems (e.g., flooding at Pensacola after Hurricane Ivan) as motivation for the theory developed, though the theoretical results are rarely brought home with identification of parameters from data. The chapters treat finite differences for solving differential equations, compartment modeling, diffusion problems, transport problems, and schemes for solving partial differential equations numerically. Each chapter has a few modeling exercises; they are open-ended and appear quite challenging.

Grinstead, Charles M., William P. Peterson, and Laurie Snell, *Probability Tales*, American Mathematical Society, 2011; ix + 237 pp, $42 (P). ISBN 978-0-8218-5261-3.

This book treats, in detail suitable for a mathematics student, four topics: streaks, the stock market, lotteries, and fingerprints, with the first two taking up one-third of the book each. The spirit is similar to Grinstead and Snell's *Introduction to Probability* (2nd ed., 1997), but with the relaxed atmosphere of being able to treat a topic at length and to assume some modest probability background on the part of the reader. [Thanks to Bruce Atwell.]

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

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS