π in the Sky

The Pacific Institute for the Mathematical Sciences

http://www.pims.math.ca/pi

March 2003

ASTech RECIPIENT 2002

π in the Sky is a semi-annual publication of

The Pacific Institute for the Mathematical Sciences

PIMS is supported by the Natural Sciences and Engineering Research Council of Canada, the British Columbia Information, Science and Technology Agency, the Alberta Ministry of Innovation and Science, Simon Fraser University, the University of Alberta, the University of British Columbia, the University of Calgary, the University of Victoria, the University of Washington, the University of Northern British Columbia, and the University of Lethbridge.

This journal is devoted to cultivating mathematical reasoning and problem-solving skills and preparing students to face the challenges of the high-technology era.

Editor in Chief

Nassif Ghoussoub (University of British Columbia)
Tel: (604) 822–3922 E-mail: director@pims.math.ca

Editorial Board

John Bowman (University of Alberta)
Tel: (780) 492–0532 E-mail: bowman@math.ualberta.ca
Florin Diacu (University of Victoria)
Tel: (250) 721–6330 E-mail: diacu@math.uvic.ca
Giseon Heo (University of Alberta)
Tel: (780) 492–8220 E-mail: gheo@ualberta.ca
Klaus Hoechsmann (University of British Columbia)
Tel: (604) 822–5458 E-mail: hoek@pims.math.ca
Dragos Hrimiuc (University of Alberta)
Tel: (780) 492–3532 E-mail: hrimiuc@math.ualberta.ca
Wieslaw Krawcewicz (University of Alberta)
Tel: (780) 492–7165 E-mail: wieslawk@shaw.ca
Volker Runde (University of Alberta)
Tel: (780) 492–3526 E-mail: runde@math.ualberta.ca
Carl Schwarz (Simon Fraser University)
Tel: (604) 291–3376 E-mail: cschwarz@stat.sfu.ca

Copy Editor

Barb Krahn & Associates (11623 78 Ave, Edmonton AB)
Tel: (780) 430–1220, E-mail: barbkrahn@shaw.ca

Technical Assistant

Mande Leung (University of Alberta)
Tel: (780) 710-7279, E-mail: mtleung@ualberta.ca

Addresses:

π in the Sky	*π in the Sky*
Pacific Institute for	Pacific Institute for
the Mathematical Sciences	the Mathematical Sciences
449 Central Academic Bldg	1933 West Mall
University of Alberta	University of British Columbia
Edmonton, Alberta	Vancouver, B.C.
T6G 2G1, Canada	V6T 1Z2, Canada
Tel: (780) 492–4308	Tel: (604) 822–3922
Fax: (780) 492–1361	Fax: (604) 822–0883

E-mail: pi@pims.math.ca
http://www.pims.math.ca/pi

Contributions Welcome

π in the Sky accepts materials on any subject related to mathematics or its applications, including articles, problems, cartoons, statements, jokes, etc. Copyright of material submitted to the publisher and accepted for publication remains with the author, with the understanding that the publisher may reproduce it without royalty in print, electronic, and other forms. Submissions are subject to editorial revision.

We also welcome **Letters to the Editor** from teachers, students, parents, and anybody interested in math education (be sure to include your full name and phone number).

Cover Page: This picture was created for *π in the Sky* by Czech artist Gabriela Novakova. The scene depicted was inspired by the article by Garry J. Smith and Byron Schmuland, "Gambling with Your Future—Knowing the Probabilities," which is published on page 5. Our readers will recognize Prof. Zmodtwo, who was also featured on the cover page of the September 2002 issue. This time, Prof. Zmodtwo gets into trouble in the Royal Casino, where he tries to use math to change his odds.

CONTENTS:

Math Opinion

This column is an open forum. We welcome opinions on all mathematical issues: research, education, and communication. Please feel free to write.

Opinions expressed in this magazine do not necessarily reflect those of the Editorial Board, PIMS, or its sponsors.

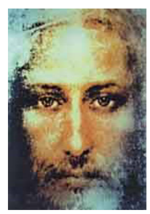
Don't be too quick to apply quantitative models to human population history*

by

Thomas Hillen†

Recent issues of *π in the Sky* featured a couple of interesting articles that apply mathematical modelling to human history [3, 7, 9] to illustrate apparent contradictions in the established sequencing of historic events. This was taken as far as the implication of Kasparov [7] and Fomenko [1, 2] that for the last 3000 years of human history, about 1000 years have been artificially added.

I took a look at Fomenko's books to get a better understanding of what he is doing and how his argument is justified. As I started reading, I felt like I had dynamite in my hands. Let me briefly describe what he does: Fomenko uses statistical methods to compare historical texts like chronicles and annals. He assumes that important and outstanding events received more attention in chronicles than boring events. Hence there must be more text available about important years and less text about not so important years. With this assumption, it is possible to "map" a historical period according to the relative importance of the years, which gives a kind of historical fingerprint (volume graph in Fomenko's book). His hypothesis is: if two independent chronicles show comparable volume graphs, then they are most likely related, or, depending on the strength of the correlation, they describe the same events. Fomenko uses a huge amount of data to verify his hypothesis and to analyze historical data. He comes to the fantastic conclusion that the known history of the last 3000 years consists of four copies of one "true" history with a total length of about 1700 years. Which, in fact, suggests that the Roman empires (first, second, third, and holy) are really just four copies of the one Roman empire that really existed. There are four copies of Julius Caesar in the history and even four copies of Jesus Christ, one of whom (he suggests) was Gregory VII Hildebrand, who lived in the 11th century AD. Fomenko's books contain more of these statistical parallels, which do not necessarily prove a new "truth," but which certainly justify questioning classical history and stimulate discussion.



Jesus Christ

Gregory VII Hildebrand

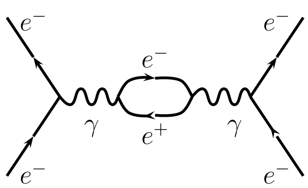
There are other articles that question classical history, such as the one written by J. Kessler [8]. In the end, it was his article that motivated me to write something about population models and their use in science. Kessler stipulates a linear dependence between a civilization event (like the invention of fire, language, printing, or the Internet) and its corresponding period of realization, which is the time needed to get the new invention established in a population. He then uses his "model" to argue against established history, which, in my opinion, goes way too far! We cannot expect that a simple model like this really describes human history. It might elucidate certain relationships, but this model certainly has its limitations.

Let us now look at the modelling process in general. Kasparov uses an exponential growth model to support his argument, Fomenko uses stochastic analysis and statistical methods, and Kessler uses a linear function. The models of Kasparov and Kessler are similar in that they are deterministic models used to describe the development of a population and to make predictions about its future. Fomenko's model is different in that Fomenko does not attempt to describe human development; he uses statistical methods to analyze the data produced by "real" history, as recorded in historical texts. It is clear, however, that each model, statistical or deterministic, has its limitations. There are always situations where a chosen model is not applicable. In a companion article in this issue, I present a deterministic model, the Verhulst model of population growth [5]. I will show how it is successfully used to describe cell growth in a petri dish, but I will also discuss the limitations of the model and even contradictory predictions, if the present model is not used appropriately.

* The "Be Careful with that Axe" illustration was created by Martin Hongsermeier. We publish it here with the artist's permission.

† **Thomas Hillen** is a professor in the Department of Mathematical Sciences at the University of Alberta.
His web site is http://www.math.ualberta.ca/~thillen/ and his E-mail address is thillen@math.ualberta.ca.

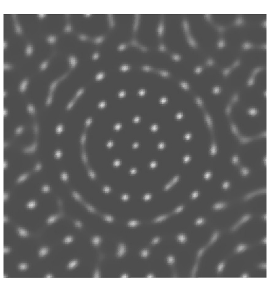
The success of mathematical modelling stems from theoretical physics. Newtonian mechanics is completely formulated in mathematical terms, and it proves immensely useful for quantitative descriptions of (macroscopic) moving objects. But still, there are many examples where Newtonian mechanics is not applicable. For small objects, it is replaced by quantum mechanics, and for large or fast objects, the Newtonian theory must be extended to Einstein's theory of relativity. These theories provide descriptions of physical observations with high accuracy. For that reason, they are sometimes called "laws"(i.e., Newton's law, Coulomb's law, Maxwell's law, Fourier's law, etc.) People tend to forget that these "laws" are not unbreakable natural laws, but rather "models" for nature. And of course, models always have limitations. Consider the never-ending invention of new particles on the subatomic level: quarks, leptons, gluons, or antiparticles. In my understanding, these are not really particles; they are "models" for observed energy relations. If a physicist says "quark" for example, he means a model of an object that is characterized by certain quantum numbers [4], like spin 1/2, Baryon number 1/3, Lepton number 0 and charge +2/3 or −1/3.



Feynman diagram

Furthermore, a "particle" is usually represented by a solution of a quantum dynamical Schrodinger equation (or generalizations), which again is a model for electromagnetic interactions. In my understanding, a question like "Does this new particle exist?" must be understood as, "Does this model describe some experiment that cannot be described without this model?"

My field of research is biomathematics, and I use models to describe movement and pattern formation in cell population [6]. Some models show a very good agreement with experimental data. They are, moreover, well suited to identifying basic principles that allow a comparison of slime molds to leukocytes, or even finding parallels with cells in an embryo. In reality, however, a cell population is a much more complex system than a physical system. Moreover, the physical world is an intrinsic part of the cell population. Hence, we can never expect to model a cell population down to the physical properties of its underlying molecules and proteins, etc. Even with modern computational power this is an impossible enterprise. What we can do, however, is to work in layers: model the microscopic events first, and then use scaling and homogenization to derive macroscopic models.



Numerical simulation of bacterial pattern formation (with Y. Dolak)

When it comes to the modelling of human populations, we face the fact that a population usually consists of states, towns, and tribes, which consist of many individuals, each of whom has many cells and molecules, etc. Hence, we can expect that it is not easy to model a population as a whole, in particular using simple models like exponential growth or linear dependence, which are most likely not realistic. But it is not impossible to work with macroscopic models for populations. They are, for example, successfully used to understand epidemic spread. Models for HIV transmission, for example, contributed to the development of prevention strategies and control mechanisms. But again, the modeler has to be very careful and has to know the model limitations. While the "laws" of physics appear to be universal inside their field of applicability, population models are flexible and they can be adapted as soon as some new information is available. This is an important difference. Mathematical population models should not be used as "laws" that are equal to "truth." This could certainly lead to the misuse of quantitative modelling.

Let me return to the discussion of human chronology. I think that mathematical modelling is not needed to question the standard historical scale. The counterexamples and open questions formulated by Kasparov, Kessler, Fomenko, and others should provide sufficient reason to reinvestigate historical events. To get a complete picture of the "true" historical chronology, one possibility is to follow these three steps: (i) use physical methods to mark astronomical events, like supernovae, comets, and solar and lunar eclipses; then (ii) identify historical events that correlate with these astronomical events; and finally (iii) relate other dates to the events of part (ii). Keep in mind all the open questions, and write the chronology without political or religious intentions.

So, why do I say, "Be careful with that axe, Eugene"? The "axe" is mathematical modelling and "Eugene" is everyone who intends to apply this tool to human history. Be careful, or you will hurt yourselves!

References:

[1] A.T. Fomenko. Empirico–Statistical Analysis of Narrative Materials and Its Applications to Historical Dating, 1: The Development of the Statistical Tools. Kluwer Academic Press, Dordrecht, (1994).

[2] A.T. Fomenko. Empirico–Statistical Analysis of Narrative Materials and Its Applications to Historical Dating, 2: The Analysis of Ancient and Medieval Records. Kluwer Academic Press, Dordrecht, (1994).

[3] N. Ghoussoub and K. Hoechsmann. Mathematics: A Tool for Questioning. *π in the Sky*, September, 9–11, (2002).

[4] F. Halzen and A.D. Martin. Quarks and Leptons: An Introductory Course in Modern Particle Physics. Wiley, New York, (1984).

[5] T. Hillen. Applications and Limitations of the Verhulst Model for Populations. *π in the Sky*, December, (2002).

[6] T. Hillen. Hyperbolic Models for Chemosensitive Movement. Math. Models Methods Appl. Sci., 12(7), (2002).

[7] G. Kasparov. Mathematics of the Past. *π in the Sky*, September, 5–8, (2002).

[8] J. Kessler. Civilizing Events and Chronology. In: 2nd International Meeting: A Revised Chronology and Alternative History, Rüspe, Germany (June 2001).

[9] W. Krawcewicz. Decoding Dates from Ancient Horoscopes. *π in the Sky*, September, 12–15, (2002).

Gambling With Your Future—Knowing the Probabilities

Garry J. Smith†
and Byron Schmuland*

Gambling is the wagering of valuables on events of uncertain outcome. This definition implies that an element of risk is involved and that there is a winner and a loser—money, property, or other items of value change hands. Gambling also implies that at least two parties are involved—a person cannot gamble alone, and the decision to wager is made consciously, deliberately, and voluntarily.

In everyday language, the word "gambling" has broad currency; for example, activities such as farming, fishing, drilling for oil, marriage, or even crossing a busy street are sometimes referred to as gambles. When used in this imprecise fashion, the concept of risk is confused with the notion of a gamble; the main distinction being that the aforementioned activities are not "games of chance" organized specifically to induce wagering. Certain gray areas such as speculative investments and playing the stock market may or may not be construed as gambling, depending on the context and circumstances.

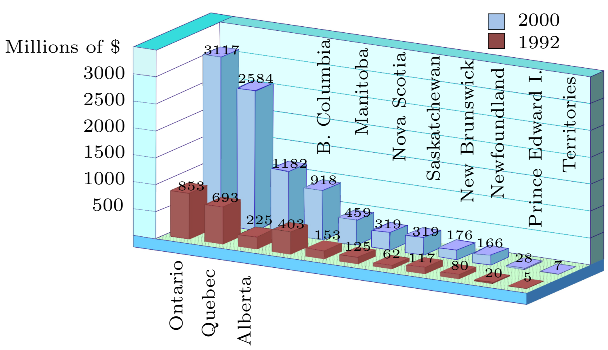


Figure 1

Total revenue wagered on lotteries, casinos, and video lottery terminals in Canada, minus prizes and winnings.

Governments and the gambling industry prefer the word "gaming," a euphemism for gambling designed to soften public perception of an activity that may evoke images of illegal activities engaged in by unsavory characters. Widespread use of the term "gaming" is intended to recognize and reinforce the activity's now legal and more acceptable status.

† **Professor Garry J. Smith** is Gambling Research Specialist with the Faculty of Extension at the University of Alberta. He has been investigating gambling issues for over 15 years. His E-mail address is garry.j.smith@ualberta.ca.

* **Byron Schmuland** is a professor in the Department of Mathematical and Statistical Sciences at the University of Alberta. His E-mail address is schmu@stat.ualberta.ca.

Until several decades ago, Canadian legal gambling was limited to horse racing and games of chance on summer fair midways. This situation changed as a result of amendments to the Criminal Code of Canada in 1969 (which allowed lotteries, charity bingos, and casinos) and 1985 (which legalized electronic gambling formats such as slot machines and video lottery terminals). Legal gambling in Canada now operates on a scale that was unimagineable 30 years ago, not only because of the proliferation of new games and gambling outlets, but also because of relaxed provincial regulations that permit gambling venues to be open longer hours and seven days a week, increased betting limits, gaming machines equipped with note acceptors, and on-site cash machines.

Legalized gambling in Canada has become a huge commercial enterprise. Provincial governments have capitalized on citizens' growing tolerance toward a previously frowned-upon social vice to fill their coffers. Figure 1 and 2 show recent Statistics Canada (2002) data attesting to the pervasiveness of legal gambling in Canada and the economic importance of gambling revenues to Canadian provinces.

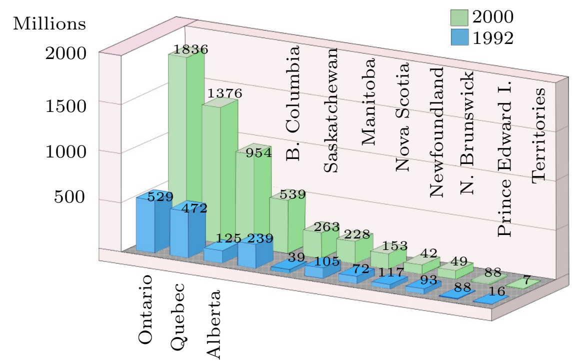


Figure 2

Net income of provincial governments from total gambling revenue less operating costs.

While overall gambling revenue in Canada expanded nearly four-fold between 1992 and 2000, the increase was mainly due to the popularity of casinos, lotteries, video lottery terminals, and slot machines; their share of total Canadian gambling revenues is 31 percent, 28 percent, 25 percent, and 15 percent respectively. Horse racing makes up less than one percent.

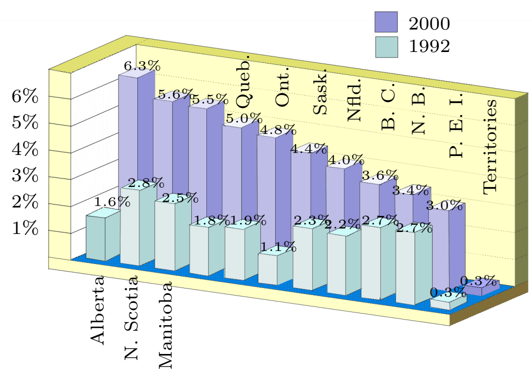


Figure 3

Share of total revenue (based on 2000 gambling revenue and 1999 total provincial revenue).

For most participants, engaging in gambling is a harmless leisure pastime; that is, the games are played infrequently and players adhere to preset time and spending limits. A small percentage (three to four percent) of adult Canadians have difficulty controlling their gambling behavior; that is, they habitually wager more than they can afford to lose and their excessive gambling jeopardizes personal relationships, job and school productivity, as well as mental and physical well-being.

Gambling formats differ in their addictive potency, depending on whether they are classified as "continuous" or "non-continuous" games. Continuous games are seen as inherently more exciting because rapid-fire sequences of wager, play, and outcome are possible within a short time span; examples include video lottery terminals, slot machines, and most casino games. Non-continuous games include lotteries and raffles, where the sequence of wager, play, and outcome may be spaced out over days or weeks. Problem gamblers invariably gravitate to continuous games.

Legalized gambling has become a huge commercial enterprise. Governments have capitalized on citizens' growing tolerance toward a previously frowned-upon social vice.

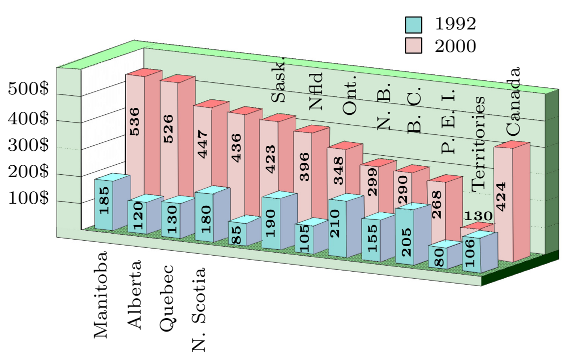


Figure 4

Expenditure per capita (amount spent on gambling by Canadians 18 years of age and older in 1992 and 2000).

Controlled gambling differs from problem gambling in the following ways:

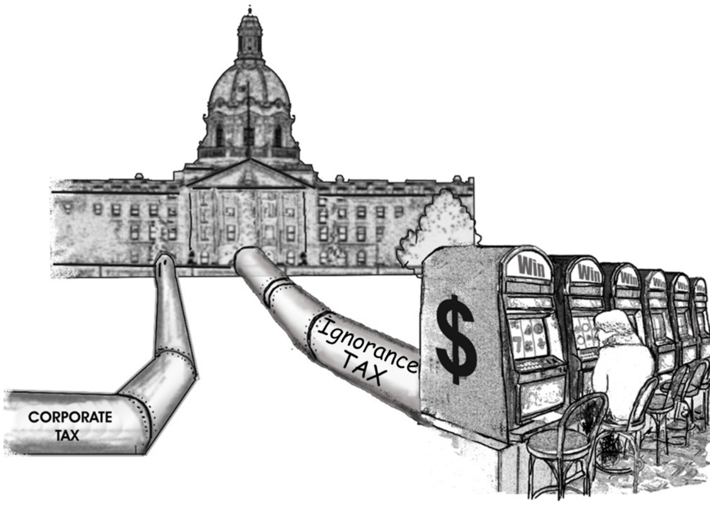
- Problem gamblers frequently "chase" their losses; that is, return as soon as possible to try and win their money back. Controlled gamblers are philosophical about losses; they see it as an entertainment fee and do not feel compelled to recoup lost funds.
- Controlled gamblers are more likely to gamble for social reasons or entertainment; the main motivations for problem gamblers are gambling to win money, to get an adrenalin rush, or to escape from boredom or loneliness.
- Problem gamblers are prone to using gambling to reach an altered state of reality; for example, to take on another identity or enter a trance. This is because they are usually unhappy with themselves and/or their life circumstances; gambling is a temporary haven where they can escape the dreariness of their lives.
- Controlled gamblers generally look forward to playing, enjoy the action, and in retrospect, have a positive feeling about their involvement. Problem gamblers also look forward to the action, but alternate between extremes of excitement and depression while playing. Afterwards, problem gamblers feel guilty and disconsolate, usually because they have lost more than they can afford.
- Controlled gamblers are more inclined than problem gamblers to be aware of the odds or payout percentages of various games and to prefer gambling formats where the application of skill can influence the outcome. Conversely, problem gamblers are more likely to fall prey to what psychologists call "irrational beliefs," "erroneous perceptions," the "illusion of control," and "magical thinking." Examples of these faulty cognitions in a gambling context are as follows:

Irrational beliefs result from a lack of understanding of probability theory; for example, a typical gambling fallacy is to believe in the so-called "law of averages." If, in playing roulette, a red number turns up ten times in a row, there is a tendency for the uninformed gambler to think that the next number is bound to be black, the rationale being that red or black is a 50/50 proposition and, given the fact that red has come up 10 times in a row, the law of averages would indicate that a black number is due. The gambler fails to recognize that each spin of the roulette wheel is an independent event; the odds are still 50/50 no matter what occurred on the last play, the last dozen plays, or whenever.

Erroneous perceptions are similar to irrational beliefs in that the gambler selectively responds to information that he/she thinks has a bearing on the outcome of a wager. For example, problem gambling slot machine or video lottery terminal players are known to be misled by the "near miss" phenomenon. This can occur when a slot machine player almost hits a jackpot; that is, the winning sequence of symbols is only one removed from the centerline but plainly visible to the player. This seeming "near miss" may induce some players to believe mistakenly that a big win is just around the corner. What the player fails to appreciate is that the machines are operated by a randomly programmed microchip and that all misses are equal, no one spin is any nearer than another to winning the prize. Some experts contend that the "near miss" feature is intentional, programmed into the machines to keep gamblers playing longer.

The **illusion of control** is a concept that refers to a gambler's belief in his/her ability to influence the outcome of a wager, when, in fact, no control is possible. For example, some gamblers believe that there are systems of play that will allow them to beat slot machines or win the lottery, even though for these activities there is no optimal playing strategy; they are strictly random outcomes. The illusion of control can also apply to gambling formats that feature an element of skill such as betting on the horses or sports events, because gamblers either tend to overestimate their own skill or discount the fact that luck is an important outcome determinant in these games as well.

Magical thinking is a belief in a thing, object, or action that is not in conformity with scientific knowledge; in other words, an unjustified or misdirected belief. In a gambling context, magical thinking is evidenced in lottery ticket buyers who regularly play their "lucky" numbers, even though a lottery draw is a pure chance event; or in bingo players who carry their "lucky charms," believe in wearing their "lucky coats," or sitting in their "lucky chairs." Players are convinced of the "special powers" of these objects and mistakenly believe that they increase their chances of winning.



What are the real gambling hold percentages and odds, and how are they calculated?

By definition, all forms of gambling contain an element of risk. Consistent winning at any gambling format is extremely rare and only possible if a player meets all of the following criteria:

1. Participates in games where there is a high skill component (e.g., poker, bridge, pool, golf, horse racing) and the player has the requisite skills and a thorough knowledge of the game's odds and strategies;
2. Has a tightly controlled system coupled with a high tolerance for drudgery and patience to wait for optimal betting opportunities;
3. Has strong emotional control to ride out the inevitable losing streaks without doubting their system or abilities, or succumbing to the irrational thinking patterns described above; and
4. Takes a strong-willed approach to money management; that is, knows when to bet and how much and knows when not to bet.

While precious few bettors possess all these attributes, one thing to be learned from this disciplined approach is being able to recognize favourable and unfavourable gambling situations. Table 1 lists hold percentages for the major legal gambling formats in Canada. The "hold percentage" is the difference between the total amount wagered and the amount of money returned in winnings; in other words, the amount retained by the gambling operator. (It is important to note that not all of the hold percentage is profit; a portion is used to pay the overhead expenses required to run the games).

Based on the hold percentages shown in Table 1, it is obvious that most of these gambling formats are heavily weighted against the player; the chances of coming out ahead in the long run are negligible. Raffles, lotteries, bingo, video lottery terminals, and pull tickets are pure chance forms of gambling where no skill can be applied to improve one's chances of winning. There is an element of skill involved in Sport Select games, but any skill involved is negated by the fact that to win, a player has to win multiple games and win them all; in addition, tie games represent a third possible game outcome that diminishes a player's odds of winning. "Casino games" is a misleading category because some games are pure chance events (e.g., roulette, baccarat, and slot machines), while others require some skill (e.g., poker, blackjack), and for each game, there are a wide variety of wagering options, each with their own odds and payback structure. For example, some bets are long shots (e.g., picking a specific number in roulette), versus others that give the player almost a 50 percent chance of winning (e.g., betting red or black or odd or even in roulette). Horse racing offers the most favourable payback percentage to consumers, but even there, the hefty 19 percent "house edge" is difficult for knowledgeable bettors to overcome.

Based on the hold percentages, it is obvious that most of these gambling formats are heavily weighted against the player; the chances of coming out ahead in the long run are negligible.

Estimated Hold Percentages for Various Canadian Legal Gambling Formats

Horse Racing	19%
Casino Games	21%
Pull Tickets	26%
Video Lottery Terminals	30%
Bingo	35%
Sport Select	37%
Lotteries	55%
Raffles	57%

Table 1

Outlined below are the odds for various gambling formats and how they are derived.

A roulette wheel

A roulette wheel has 38 numbered slots: 18 red, 18 black, and two green. You can bet on an outcome as simple as the colour or number that appears, or make more complicated bets based on groups of numbers (e.g., split, street, corner, double street). The chance of winning a single bet is simply the ratio of favourable outcomes to total outcomes. For instance, if you bet on "black," then the chance of winning is $18/38 = 0.473684$. This is not quite a fair bet, but the payout is 1 to 1, as if it were fair. This means the house edge on colour bets is created by those innocent-looking green slots.

Craps is a two-stage game that uses a pair of dice. Three things can happen on the first roll: you win immediately (7 or 11), you lose immediately (2,3, or 12), or you establish a point (any other value). If the first roll establishes a point, you continue to roll the dice until either you roll your point again (win) or you roll a 7 (lose). The chance of winning at craps is the combination of an immediate win on the first roll, or hitting your point before rolling a 7. A rather complicated calculation puts the chance of winning at 244/495 = 0.492929. You would do better in craps with a "don't pass" bet. In this case, you are betting against the roller, except that a 12 on the first roll counts as a tie. Subtracting the outcomes that give a tie, we find that the chance of winning with a "don't pass" bet is effectively (251/495) − (1/36) divided by 1 − (1/36); that is, 0.492987.

A Craps table

A useful rule of thumb is that the more skill needed to play a game, the better your odds of winning. Slot machines and lotteries give you the worst odds; you would do better with roulette. As we calculated, the game of craps gives even better odds than roulette. Games that are not purely random, but require skill, like poker or blackjack, give the best possible gambling odds (if played skillfully!).

The most popular lottery in Canada is Lotto 6–49. Six numbers are randomly chosen from 1 to 49; the prize depends on how many of these match the numbers on the ticket. If three numbers match you win $10; if all six numbers match you win the jackpot. Of course, there are other prizes for matching four or five numbers as well. What are your chances?

The number of possible ticket combinations is $\binom{49}{6}$ = 13 983 816. Your chance of winning the jackpot is therefore one out of 13 983 816, which is 7.15×10^{-8}.

As for matching three numbers, consider the numbers from 1 to 49 divided into two groups: the six numbers on your ticket, and the 43 numbers that aren't on your ticket. To win $10, you need exactly three from the first group, and three from the second group. The number of Lotto 6–49 drawings of that type is

$$\binom{6}{3} \times \binom{43}{3} = \frac{(6)(5)(4)}{(3)(2)(1)} \times \frac{(43)(42)(41)}{(3)(2)(1)}$$
$$= 20 \times 12341 = 246820.$$

Thus, the chance of matching exactly three numbers is 246820/13983816 = 0.017650.

We can find all the Lotto 6–49 probabilities in the same way. The bottom of the ratio is always equal to the total number of Lotto 6–49 draws: $\binom{49}{6}$. The top of the ratio always has two terms: 43 choose something times six choose something. The term with 43 represents the number of ways to choose from the 43 values not on your ticket, and the other term represents the number of ways to choose from the six values on your ticket. If you think of the numbers as "good" or "bad" according to whether or not they are on your ticket, then $\binom{43}{6}\binom{6}{0}$ is the number of draws that result in six bad numbers and zero good numbers. Similarly, $\binom{43}{5}\binom{6}{1}$ is five bad numbers and one good number, and so on. The following table gives the complete lowdown on Lotto 6–49.

Matches	Probability
0	$\frac{\binom{43}{6}\binom{6}{0}}{\binom{49}{6}} = 0.43596\,49755$
1	$\frac{\binom{43}{5}\binom{6}{1}}{\binom{49}{6}} = 0.41301\,94505$
2	$\frac{\binom{43}{4}\binom{6}{2}}{\binom{49}{6}} = 0.13237\,80290$
3	$\frac{\binom{43}{3}\binom{6}{3}}{\binom{49}{6}} = 0.01765\,04039$
4	$\frac{\binom{43}{2}\binom{6}{4}}{\binom{49}{6}} = 0.00096\,86197$
5	$\frac{\binom{43}{1}\binom{6}{5}}{\binom{49}{6}} = 0.00001\,84499$
6	$\frac{\binom{43}{0}\binom{6}{6}}{\binom{49}{6}} = 0.00000\,00715$

Adding the first three probabilities in the table shows that there is a better than 98 percent chance of losing your dollar. The odds of winning $10 (matching three numbers) is $0.01765 \approx 1/56$, so on average you spend $56 to win $10. A last bit of Lotto 6–49 trivia: if you play twice a week, every week for 1000 years, the chances are better than 99 percent that you will never, ever win the jackpot!

Gambling can be fun if treated as entertainment, and when done in moderation. However, it is a bad way to invest your money or to try to get rich. The casino (or the government, in the case of lotteries) uses mathematical probability to ensure it retains the edge needed to guarantee profits. There are no games that favour the player and there is no legal betting system, no matter how complex, which will alter this basic fact. They rely on your ignorance to line their pockets.

References

[1] Devereux, E. (1979). Gambling. In The International Encyclopedia of the Social Sciences. Vol. 17. New York: Macmillan.

[2] Langer, E. (1975). The Illusion of Control. Journal of Personality and Social Psychology, 32, 311–328.

[3] Scarne, J. (1975). Scarne's New Complete Guide to Gambling. London: Constable.

[4] Smith, G. (1990). Pools, Parlays, and Point Spreads: A Sociological Consideration of the Legalization of Sports Gambling. Sociology of Sport Journal. 7, 271–286.

[5] Smith, G., Volberg, R. & Wynne, H. (1996). Leisure Behavior on the Edge: Differences Between Controlled and Uncontrolled Gambling Practices. Society and Leisure, 17(1), 233–248.

[6] Statistics Canada, (2002). Taking a Chance? Perspectives, Autumn.

[7] Walker, M. (1995). The Psychology of Gambling. Oxford, UK: Butterworth-Heinemann.

On the Dynamics of Karate

by Florin Diacu†

The origins of martial arts can be traced back to ancient times and the systems of self-defense and fighting designed by Oriental priests and Asian warriors. Derived from those systems, karate, meaning "empty hand," was developed in Okinawa in the early 17th century after the Japanese conquered the island and banned the use of all weapons. Today, millions of people are practising karate all over the world. There exist many karate styles, four of which are recognized by the World Karate Federation: goju, shito, shotokan, and wado. No style is superior to any other. All of them lead to similar results, but each of them follows specific ideas and reflects a different philosophy.

Picture 1

Sally Chaster, 1st dan black-belt and chief instructor at the Kuwakai club in Victoria, executes a forward punch (junzuki).

† **Florin Diacu** is Professor of Mathematics and the Director of the Pacific Institute for the Mathematical Sciences at the University of Victoria. He practices wado-style karate with the Kuwakai Club in Victoria. Some popular science articles on other subjects, from celestial mechanics to financial markets, are posted on his web site: http://www.math.uvic.ca/faculty/diacu/index.html. His E-mail address is diacu@math.uvic.ca.

Our goal here is to use a few simple mathematics and physics tools to analyze the dynamics of karate and to draw several conclusions on the efficiency of various techniques. Let us start with taking a look at the forward punch (junzuki, see Picture 1). When performing a junzuki, the goal of the karate practitioner (karateka) is to keep the body in balance and achieve maximum energy when the knuckles hit the target. The fist travels a straight distance and rotates by approximately 180 degrees. Assuming that the rotation is uniform and that the fist and the forearm are approximated with a cylinder of radius r, the energy E is given by the formula:

$$E = mgh + \frac{1}{2}mv^2 + \frac{1}{2}mr^2\omega^2,$$

where m is the mass used in the punch, h is the difference in the height of the body from the initial position to the position when the punch hits the target (when stepping forward, the body drops 15 to 20 cm), $g = 9.8$ m/s^2 represents the gravitational acceleration, v denotes the velocity of the fist, and ω is the angular velocity of the fist's rotation. The three terms appearing on the right side of this equation are called: potential, kinetic, and rotational energy, respectively.

This formula allows us to draw several conclusions.

1. The greater the mass, the higher the energy. We see that the energy grows linearly with the mass. This implies that if X is twice as heavy as Y, then X converts two times more energy than Y. Apparently, we cannot do much about this quantity, which depends on the frame of the body. An arm usually weighs about 10 percent of a person's total body, but we can increase the mass of a punch by stepping forward. However, unlike street fighters, who often engage most of their body mass in a punch at the expense of losing their balance, the karateka chooses to use less mass to favour stability. As we will see below, there are better ways to increase the energy of a punch or a kick without losing balance and becoming vulnerable to a counterattack.

2. The lower the drop, the higher the energy. The above formula shows that the energy grows linearly with the difference in height, h, when the body is dropped. The potential energy, mgh, is a substantial source since it uses the entire mass of the body. The importance of this component will become clear in the numerical example given in remark 4.

3. The higher the speed, the higher the energy. Unlike mass and height difference, which are linear quantities, speed influences energy quadratically. This means that if X and Y have the same mass but X is twice as fast as Y, then X will produce four times more energy; if X is three times faster than Y, X will produce nine times more energy. This shows that speed is an essential component in karate and in any other physical fighting game. Fast punches and kicks are not important only because they surprise the opponent, but also because of their efficiency in producing energy. A karateka who breaks boards and bricks manages to achieve the highest speed at the moment of impact.

To better appreciate the importance of speed, let us note that some simple computations show the following facts:

- If X weighs 50 kg and Y weighs 70 kg, then X must punch only 18 percent faster to achieve the same effect as Y;
- If X weighs 50 kg and Y weighs 100 kg, then X must punch only 41 percent faster to achieve the same effect as Y.

This shows that women, who are in general smaller than men, can deliver equally effective punches if they increase their speed.

4. The effect of the fist's rotation is negligible. Contrary to what most people think, the effect of the fist's rotation is negligible in a punch. The best way to see this is through a numerical example.

Suppose that a karateka weighing 70 kg performs a forward punch (junzuki). Assume that the mass involved in the punch is that of the arm alone (approximately 7 kg). The average speed achieved by a black-belt karateka's fist at the moment of impact is about 7 m/s (see the table below) and that of the first rotation, ω, is about 5π rad/s (i.e., the fist rotates 180 degrees in 0.2 seconds). Let us also assume that the radius r of the cylinder that approximates the fist and the forearm is 3 cm = 0.03 m. The drop in height is approximately 20 cm = 0.2 m. Then the potential energy, E_P, kinetic energy, E_K, and rotational energy, E_R, take the following values, measured in Joule (J) (recall that 1 J = 0.239 calories):

$$E_P = mgh = 70 \times 9.8 \times 0.2 = 137.2 \text{ J},$$

$$E_K = \frac{1}{2}mv^2 = \frac{1}{2} \times 7 \times 7^2 = 171.5 \text{ J},$$

$$E_R = m(r\omega)^2 = \frac{1}{2} \times 7 \times (0.03 \times 5\pi)^2 = 0.78 \text{ J}.$$

This shows that the rotation accounts for 0.45% of the kinetic energy, 0.57% of the potential energy, and only 0.25% of the total energy. One quarter of a percentage point is a negligible quantity. However, we can see that the energies converted by the drop in height and the motion of the arm are comparable. This also explains the principle of keeping the body at the same height in order to conserve energy. Every up-and-down move by only 20 cm uses almost as much energy as a punch at 7 m/s.

5. The longer the distance, the higher the energy. We will now show that the energy changes linearly with the distance the fist travels from the time of initiating the punch to the time of impact. For this, recall the following two physics formulas:

$$v = at \quad \text{and} \quad L = \frac{1}{2}at^2,$$

which indicate that the velocity, v, equals the acceleration, a, times the time, t, and that the length, L, equals half the acceleration times the square of the time. Eliminating t from the two formulas, we obtain

$$v = \sqrt{2La},$$

which means that the speed increases with the square root of the distance. Substituting v into the expression of the kinetic energy, it follows that

$$E_K = mLa.$$

This proves the linear dependence of the energy on the distance, and shows that a longer arm can reach a higher speed at impact. However, there is a drawback to this, which we will discuss next.

6. The longer the distance, the longer the time. Intuitively, this should be clear to anybody. However, the linear dependence is not. Eliminating the acceleration from the two formulas written before, we obtain

$$t = \frac{2L}{v}.$$

This means that a shorter arm will reach the target linearly faster. In other words, if X's fist travels half the distance of Y's, Y will generate two times more kinetic energy but will need double the time to reach the target. In practice, this is not entirely true since v is not constant. The speed versus the position looks like the curve in Figure 1, as it is experimentally shown in [1]. However, the linear dependence between time and length is valid.

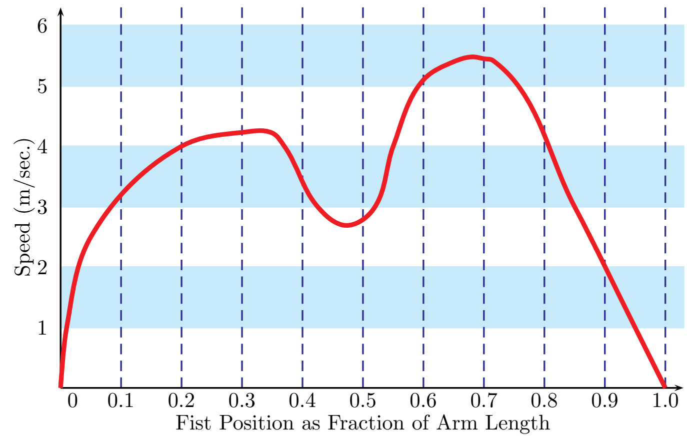


Figure 1

The speed of a fist in a forward punch as a function of its position. Data taken from a high-speed movie by J.D. Walker of Cleveland State University.

Except for the result about rotation in remark 4, these conclusions are also true in the case of kicks (e.g., the front kick, or maegeri, see Picture 2). Kicks, however, are more efficient than punches, not only because of the greater mass of the leg, which can reach up to 20 percent of a person's body, but especially due to the higher speed of the kick.

A comparative experimental study for the speeds of different techniques was done in [2]. The following table summarizes the conclusions obtained by the authors.

Technique	Max. speed
Front forward punch (junzuki)	5.7 – 9.8 m/s
Downward hammerfist block (otoshiuke)	10 – 14 m/s
Downward knife hand strike (shutouke)	10 – 14 m/s
Front kick (maegeri)	9.9 – 14.4 m/s
Side kick (yokogeri)	9.9 – 14.4 m/s
Roundhouse kick (mawashigeri)	9.5 – 11 m/s
Back kick (ushirogeri)	10.6 – 12 m/s

Table 1.

Speeds of different techniques.

We can now draw the following conclusion:

7. Kicks are between three and six times more powerful than punches. In the example described in remark 4, we computed the average potential, kinetic, and rotational energy of a junzuki punch performed by a black-belt karateka weighing 70 kg. We found the total energy was 309.48 J. Assuming now that the leg of the same person weighs 14 kg, that the speed of the kick is 12 m/s, and that there is no drop in height when executing the kick, we obtain that the energy developed by the technique, given by the kinetic energy alone, is

$$E = E_K = \frac{1}{2}mv^2 = \frac{1}{2} \times 14 \times 12^2 = 1008 \text{ J}.$$

This shows that the front kick is at least three times stronger than the forward punch. If the punch is executed without stepping forward and dropping the body, then its energy is 171.5 J, which is almost six times less than the value obtained for the kick.

Picture 2

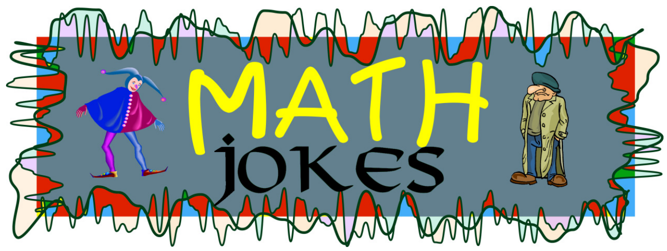
Norma Foster, 6th dan black-belt and the highest-ranked wado-karate woman in the world, performing a front kick (maegeri).

Similar estimates can be done for all kicks and punches, verifying the conclusion stated at the beginning of remark 7.

References

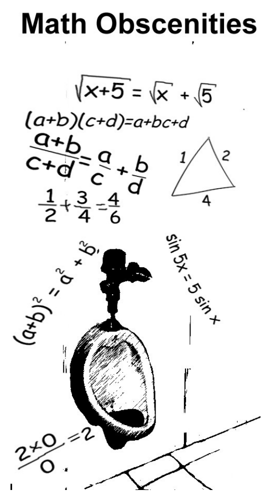
[1] Walker, J.D.: Karate Strikes, American Journal of Physics, **43** (1975), 845–849.

[2] Wilk, S.R., McNair, R.E., and Feld, M.S.: The Physics of Karate, American Journal of Physics, **51** (1983), 783–790.



A math professor, a native Texan, was asked by one of his students: "What is mathematics good for?"

He replied: "This question makes me sick! If you show someone the Grand Canyon for the first time, and he asks you, 'What's it good for?' what would you do? Well, you'd kick that guy off the cliff!"



©Copyright 2003
Wieslaw Krawcewicz

A mathematician organizes a raffle in which the prize is an infinite amount of money paid over an infinite amount of time. Of course, with the promise of such a prize, his tickets sell like hot cakes.

When the winning ticket is drawn and the jubilant winner comes to claim his prize, the mathematician explains the mode of payment: "1 dollar now, 1/2 dollar next week, 1/3 dollar the week after that . . . "

Q: What do you get if you cross an elephant with a grape?

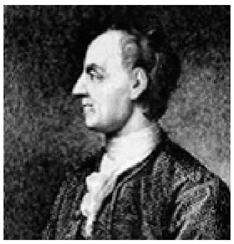
A: $|\text{elephant}||\text{grape}| \cdot \sin(\theta)$.

Theorem. Every positive integer is interesting.

Proof. Assume that this is not so; that is, there are uninteresting positive integers. Then there must be a smallest uninteresting positive integer. But being the smallest such number is an extremely interesting property!

Leonhard Euler

Alexander Litvak[†]
and Alina Litvak[*]

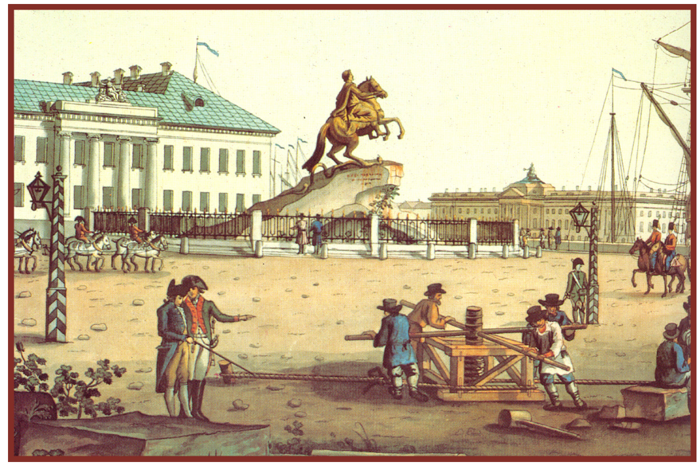


This year, St. Petersburg celebrates its 300[th] anniversary. This gorgeous city, one of the most beautiful in the world, was created by the desire and power of a single man—Russian czar Peter the Great. Peter founded the city in 1703 in an empty and swampy place. He sought to transform Russia into a more civilized, cultured, and developed country, to "westernize" it. Peter was changing Russia, almost rebuilding, and for this he needed a new face, a new capital, the most beautiful and luxurious city in Europe. The construction of St. Petersburg was extremely difficult. Thousands of workers died from disease, the cold, and the unbearable living conditions. But Peter had no time to wait. In 1712, the capital of the Russian empire moved from Moscow to St. Petersburg. Peter wanted not only to build the city of his dreams, he wanted St. Petersburg to be the cultural and scientific center of Russia. There was a lack of educated, skilled people in Russia at the time—it's hard to believe, but not all Russian noblemen could read and write, and very few spoke foreign languages. Peter sent the youth of Russian nobility to study abroad, and invited foreign specialists in different fields to work in Russia. Many architects, sculptors, and engineers from Denmark, Holland, France, Germany, Italy, and other countries came to Russia, building and decorating its cities, creating its navy, forming its industry. When Peter died, his wife Catherine the First continued his reforms. In 1725, she established the Russian Academy of Science. As was common in those times, she invited many foreigners to work in the newly-created Academy. Many great scientists came to the young capital of the Russian Empire. Among them was one the leading mathematicians of the 18th century—Leonhard Euler.

Peter the Great

[†] **Alexander Litvak** is a professor in the Department of Mathematical Sciences at the University of Alberta. His web site is http://www.math.ualberta.ca/Litvak_A.html and his E-mail address is alexandr@math.ualberta.ca.

[*] **Alina Litvak** is a software engineer employed by Intuit Canada Limited in Edmonton.



An old painting of the monument to Peter the Great in St. Petersburg

Euler was born in Basel, Switzerland on April 15, 1707. His father was a pastor and, as was usual, he wanted his son to also become a minister. He sent Euler to the University of Basel to study theology. However, it turned out that the young man had a gift for mathematics and loved it very much. Johann Bernoulli, the Swiss mathematician, paid attention to the talented student and convinced the elder Euler to allow his son to change his specialization to mathematics. Euler continued to study at the University of Basel and finished in 1726. He published his first research paper in 1726 and his second in 1727. His work on the best arrangement of masts on a ship was submitted for the Grand Prize of the Paris Academy of Science and won second place. That was a big achievement for the young mathematician. In 1726, Euler was offered the physiology post at the Russian Academy of Science.

The Russian Academy of Science in Saint Petersburg.

He accepted the offer and arrived in St. Petersburg in 1727. D. Bernoulli and J. Hermann, who were already working at the Russian Academy, helped Euler to join the mathematics–physics division, which meant that he also became a full member of the Academy. The same year, Euler married Katherine Gzel, daughter of a Swiss painter who worked in St. Petersburg.

In 1736, Euler published the two-volume work "*Mechanica, sive motus scientia analytice exposita*," where he applied mathematical analysis methods to the problems of motion in a vacuum and in a resisting environment. This work earned him world fame. Euler developed some of the first analytical methods for the exact sciences; he started to apply differentiation and integration to physical problems. By 1740, Euler had attained a very high profile, having won the Grand Prize of the Paris Academy of Science in both 1738 and 1740. He had also written the wonderful "*Direction to Arithmetic*," which was later translated into Russian. It was the first Russian book to represent arithmetic as a mathematical science.

Frederick the Great

In 1740, after the death of the Empress Anna Ioanovna, two-month-old Ioan IV was declared Emperor of Russia. As he was too young to rule, his mother, Anna Leopoldovna, became regent. Living in Russia became dangerous, especially for foreigners, and Euler decided to accept the invitation of Frederick the Great, the King of Prussia, to work in Berlin. There, Euler was met with great respect and was given the freedom to pursue his research as he wished. However Euler didn't completely end his work for Russian Academy. He was still partially paid by Russia, and he continued to write reports for the Academy and teach young Russians who arrived in Berlin. The Russians respected him so much that when his house was destroyed by Russian troops during the Russian–Prussian war, Euler received full compensation.

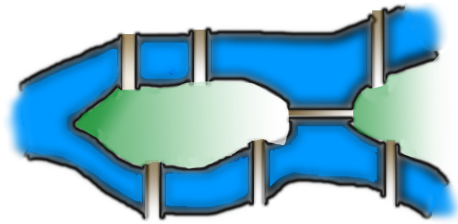
Euler's 25 years in Berlin were very busy and productive. He enjoyed great mathematical success and also found time to accomplish all kinds of social work. For example, he served on the Library and Scientific Publications Committee of the Berlin Academy and was a government advisor on state lotteries, insurance, annuities and pensions, and artillery.

Euler wrote nearly 380 articles during his Berlin period. He also wrote many scientific and popular science books, including his famous "*Letters to a Princess of Germany*," which was translated into many languages and published almost 40 times. He also led the Berlin Academy of Science after the death of Maupertuis in 1759, although he never held the formal title of President.

Euler's phenomenal ability to work is demonstrated by the fact he produced about 800 pages of text per year. That would be a significant number even for a novelist; for a mathematician, it is hardly believable. Euler made a big contribution to analysis, geometry, trigonometry, and number theory, and introduced such notation as $f(x)$ for function, $\sum$ for sum, e for the base of natural logarithm, π for the ratio of the length of a circle to its diameter, and i for imaginary unit. Euler proved the following formula for a convex polyhedron: $V + F = 2 + E$, where V is number of vertexes of the polyhedron, F is number of faces of the polyhedron, and E is number of edges of the polyhedron. This formula has the extension, very important in topology, called Euler characteristics. In addition to his work in mathematics, Euler published works in philosophy, astronomy, physics, and mechanics.

Using the graph theory that he introduced, Euler solved the following famous problem, the so-called "Königsberg's Bridges Problem."

Problem: The Pregel river in Königsberg has the form shown in the picture below. There are seven bridges across it. Would it be possible, walking through the town, to cross each bridge exactly once?



Königsberg Bridges

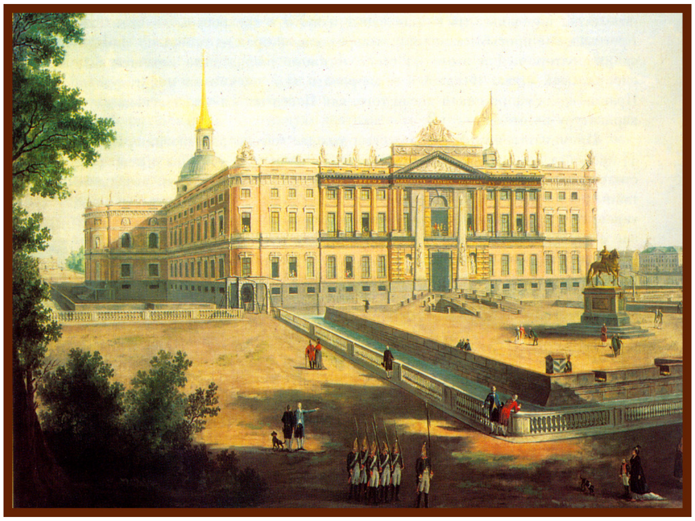
Euler was able to show that this is impossible; moreover he described precisely the form of the river and bridges required to reach an affirmative solution.

Catherine the Great

In 1762, the politics in Russia changed again. Empress Catherine II, later named "Catherine the Great," came to the throne. The atmosphere in Russian society improved dramatically. Catherine was an extraordinary person, very talented and educated. She aimed to create in Russia a regime of "*Educated Absolutism*." She invited many progressive people to Russia—she was in correspondence with Voltaire, she wrote books and plays, and she was very interested in art and in science. Catherine II also started one of the most famous and beautiful museums in the world, the Hermitage. The time of her rule is called the "Gold Century" of Russian history. Catherine the Great understood very well that the country couldn't prosper without science. She knew also that the science could enhance her prestige. She increased the budget of the Academy to 60 000 rubles per year, which was much more than the budget of the Berlin Academy.

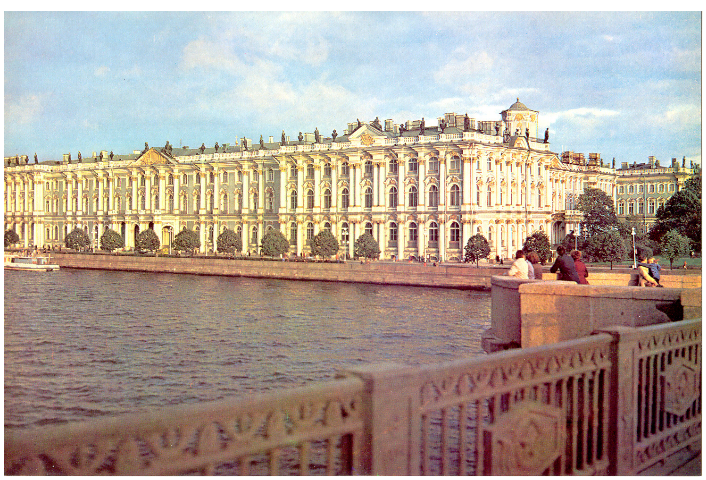
Catherine II offered Euler an important post in the mathematics department, conference-secretary of the Academy, with a big salary. She instructed her representative in Berlin to agree to his terms if he didn't like her first offer, to ensure that he would arrive in St. Petersburg as soon as possible.

In 1766, Euler returned to St. Petersburg. Soon after, he became almost blind due to a cataract in his left eye (his right eye was already very poor). However, that didn't stop him from working. Euler dictated his works to a young boy, who wrote them in German. In 1771, his home was destroyed by fire and he was able to save only himself and all of his mathematical manuscripts except the "New Theory of the Motion of the Moon." Fortunately, Euler had an exceptional memory, which helped him restore the manuscript quite quickly. After the fire, Euler was obliged to move into a new house, the interior of which was unknown to him. This was extremely difficult for a blind old man.



An old painting of the Mikhalovsky Castle in St. Petersburg

In September 1771, Euler had surgery to remove his cataract. The surgery took only three minutes and was very successful—the mathematician's vision was restored. Doctors advised Euler to avoid bright light and overloading his eyes; reading and writing were forbidden. Unfortunately, Euler didn't take care of his eyes; he continued to work and after a few days lost his vision again, this time without any hope of recovery. Euler took this quietly, with great courage. Amazingly, his productivity only increased. Despite his total blindness, Euler wrote almost half of his articles after his return to St. Petersburg.



The Winter Palace, the Emperor's residence (now the main building of the Hermitage Museum)

In 1773, Euler's wife died. They were together almost 40 years and had 13 children. At that time, the mortality rate for children was very high, and eight of their children died before reaching adulthood. After his wife's death, Euler continued to work diligently, using his son's and some of his student's eyes for reading. He worked until September 18, 1783, the last day of his life. According to his biographer, on that day Euler gave a mathematics lesson, worked on mathematics, and discussed with Lexell and Fuss the planet Uranus, recently discovered by astronomers. He died in the evening.

German and Russian postal stamps dedicated to Euler

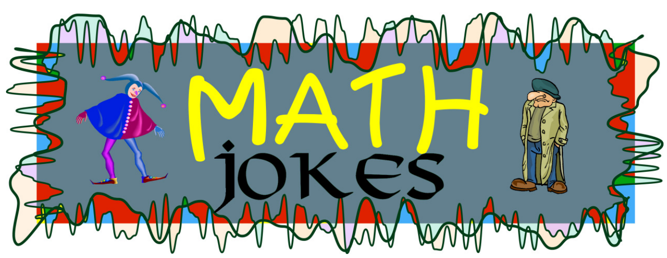
References:

[1] Leonhard Euler, J. J. O'Connor and E. F. Robertson, http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/ School of Mathematics and Statistics, University of St Andrews, Scotland.

[2] Leonhard Euler, The Brokhaus & Efron Encyclopedic Dictionary.

[3] Biography, Encyclopaedia Britannica.

[4] Mathematical Encyclopedic Dictionary, Moscow 1995.



Denis Diderot was a French philosopher in the 18th century. He traveled Europe extensively, and on his travels stopped at the Russian court in St. Petersburg. His wit and suave charm soon drew a large following among the younger nobles at the court—and so did his atheist philosophy. That worried Empress Catherine the Great very much...

Swiss mathematician Leonhard Euler was working at the Russian court at that time and, unlike Diderot, he was a devout Christian. So, the Empress asked him for help in dealing with the threat posed by Diderot.

Euler had himself introduced to Diderot as a man who had found a mathematical proof for the existence of God. With a stern face, the mathematician confronted the philosopher: "Monsieur, $(a + b^n)/n = x$ holds! Hence, God exists. What is your answer to that?"

The usually quick-witted Diderot was speechless. Laughed at by his followers, he soon returned to France.

Vedic Mathematics

by Jeganathan Sriskandarajah[†]

Sri Bharati Krsna Tirthaji

Vedic Mathematics is based on 16 sutras (or aphorisms) dealing with mathematics related to arithmetic, algebra, and geometry. These methods and ideas can be directly applied to trigonometry, plain and spherical geometry, conics, calculus (both differential and integral), and applied mathematics of various kinds. It was reconstructed from ancient Vedic texts early in the last century by Sri Bharati Krsna Tirthaji. Bharati Krsna was born in 1884 and died in 1960. He was a brilliant Indian scholar with the highest honours in the subjects of Sanskrit, Philosophy, English, Mathematics, History, and Science. When he heard about the parts of the Vedas containing mathematics, he resolved to study these scriptures and find their meaning. Between 1911 and 1918, he was able to decode from the ancient Sutras the mathematical formulae that we now call Vedic Mathematics.

The Sanskrit word Veda means knowledge, and the Vedas are considered the most sacred scripture of Hinduism referred to as sutras, meaning what was heard by or revealed to the seers. Vedas are the most ancient scriptures dealing with all branches of knowledge—spiritual and worldly. Although there is an ongoing dispute regarding the age of the Vedas, it is commonly believed that these scriptures were written at least several centuries BC. The hymns of the Rig Veda are considered the oldest and most important of the Vedas, having been composed between 1500 BC and the time of the great Bharata war, about 900 BC. The Vedas consist of a huge number of documents (there are said to be thousands of such documents in India, many of which have not yet been translated), which are shown to be highly structured, both within themselves and in relation to each other. The most holy hymns and mantras are put together into four collections called the Rig, Sama, Yajur, and Atharva Vedas. They are difficult to date, because they were passed on orally for about 1000 years before they were written down. More recent categories of Vedas include the Brahmanas, or manuals for ritual and prayer. Subjects covered in the Vedas include: grammar, astronomy, architecture, psychology, philosophy, archery, etc.

One hundred years ago, Sanskrit scholars translating the Vedic documents were surprised at the depth and breadth of knowledge contained in them. Some documents, called 'ganita sutras' (the name *'ganita'* means mathematics), were devoted to mathematical knowledge. In these sutras, which, for example, addressed the geometry of construction of sacrificial altars, geometrical figures such as straight lines, rectangles, circles and triangles are discussed in a very profound manner. There are various descriptions of the rules for transformations, including the 'Pythagorean' theorem. The proof of this theorem, as described in the Vedas, is illustrated in Figure 1.

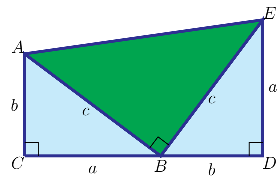


Figure 1

The area of the trapezoid $ACDE$ is equal to the sum of the areas $\triangle ABC + \triangle ABE + \triangle BDE$. Thus,

$$\frac{(a+b)^2}{2} = \frac{ab}{2} + \frac{c^2}{2} + \frac{ab}{2}$$
$$\implies c^2 = a^2 + b^2.$$

The Apollonius theorem, which states *for a triangle with sides a, b, and c and median m to the side with length a, that*

$$b^2 + c^2 = 2m^2 + \frac{a^2}{2},$$

was also described in the Vedas. Its simple proof, as presented in scripture in the Vedas, can be summarized in a few lines (see Figure 2):

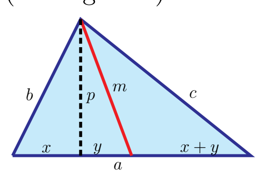


Figure 2

$$\begin{aligned} b^2 + c^2 &= x^2 + p^2 + (x+2y)^2 + p^2 \\ &= 2(y^2 + p^2) + 2(x+y)^2 \\ &= 2m^2 + \frac{a^2}{2}. \end{aligned}$$

The areas of a triangle, a parallelogram, and a trapezoid, as well as the volume of a prism, a cylinder, and a pyramid, are also discussed in sutras. The quadratic equation is utilized for the enlargement or reduction of the altar's size. The Vedic Hindus knew that the numbers $\sqrt{2}$ and $\sqrt{3}$ are irrational. Although there is no explanation in the Vedas how that was discovered, several derivations of their approximate values are embedded in the text itself. Other favourite mathematical topics in the Vedas are permutations and combinations. A special method for finding the number of combinations, called meru prastara, is described in Chandah sutras (200 BC). It is basically the same triangular array commonly known as Pascal's triangle.

The most important Hindu achievement is the decimal positional system. Let me point out that in European mathematics, the decimal system appears only after the 14th century, and the notions of subtraction and zero were not introduced until the 16th century. All of the quantities in European mathematics had dimensions and purely geometric characters.

It is amazing how advanced and sophisticated Hindu mathematics was, a thousand years before the development of European mathematics. In the Hindu decimal system, there are nine symbols called anka (which means 'mark') for the numerals from one to nine, and the zero symbol called sunya (which means 'empty'). The Hindu name for addition was Samkalita, but the terms Samkalana, Misrana, Sammelana, Praksepana, Ekikarana, Yukti etc., were also used by some writers. Subtraction was called Vyukatkalita, Vyutkalana, Sodhana, and

[†] **Jeganathan Sriskandarajah** is an instructor at Madison Area Technical College, where he recently organized the first annual "Pi Day": http://matcmadison.edu/is/as/math/mathclub/Piday03/Piday03.html. He is also the State (Wisconsin) Director for American Mathematics Competitions and a recipient of the Mathematical Association of America's Meritorious Service Award in 1998. His E-mail address is jsriskandara@matcmadison.edu.

Patana; multiplication was called Gunana, Hanana, Vedha, Ksaya; and division, regarded as the inverse of multiplication, was called Bhagahara, Bhajana, Harana. The remainder was called Sesa or Antara, and the quotient Labdhi or Labhdha. While European mathematics even in the 16th century did not consider powers of a degree higher than three (since they were not making sense from the geometric point of view), centuries earlier Hindu mathematics studied algebraic equations of degree six or higher. There was even a symbol used for the unknown, which was called Varna, and the unknown quantity was called Yavat-Tavat. Equations with one unknown were called Eka-Varna-samikarana, and equations with several unknowns were called anekavarna-samikarana. There are many more examples of advanced mathematical knowledge contained in the Vedas.

The Vedic methods in arithmetic that were discovered by Sri Bharati Krsna Tirthaji are astonishing in their simplicity. For example, multiplication of large numbers can be done in such an easy way that all the computations and the answer can usually be written in just one line. People who grasp some of the Vedic techniques sometimes dazzle audiences, pretending to be prodigies with a supernatural ability to do complicated computations quickly in their minds. However, it is important to note that no special talent is needed and anybody can take advantage of these ancient methods to improve his or her arithmetic skills. Let me emphasize that many of the mathematical methods described in the Vedas were previously unknown and created great amazement among scholars. In comparison, the circumstances surrounding the discoveries of many ancient Greek or Roman manuscripts dealing with mathematics are considered to be rather suspicious. None of these ancient manuscripts contained any "new" mathematical knowledge, previously unknown to the scientists. This is not the case for the Vedas, which continue to be analyzed, leading to new revelations.

The Vedic methods are direct, and truly extraordinary in their efficiency and simplicity. They reflect a long mathematical tradition, which produced many simplifications, shortcuts and smart tricks. Arithmetic computations cannot be obtained faster by any other known method.

Example 1. A simple idea for factorization of polynomial expressions of two or more variables is rooted in Adyamadyena Sutra—*Alternate Elimination and Retention.* Let us consider, for example, the polynomial $P(x,y,z) = 2x^2 + 6y^2 + 3z^2 + 7xy + 11yz + 7xz$, which can be factorized by setting $z = 0$:

$$P(x,y,0) = 2x^2 + 7xy + 6y^2 = (2x + 3y)(x + 2y), \qquad (1)$$

and next, setting $y = 0$:

$$P(x,0,z) = 2x^2 + 7xz + 3z^2 = (2x + z)(x + 3z). \qquad (2)$$

By comparing the obtained factorizations (1) and (2) and completing each factor with the additional terms from the other factorization, we obtain the factorization of $P(x,y,z)$:

$$P(x,y,z) = (2x + 3y + z)(x + 2y + 3z). \qquad (3)$$

Also, notice that on substituting $x = 0$, we obtain $P(0,y,z) = 6y^2 + 11yz + 3z^2 = (3y + z)(2y + 3z)$, in accordance with the factorization (3).

Example 2. It is also possible to eliminate two variables at a time. For example, consider the polynomial $Q(x,y,z) = 3x^2 + 7xy + 2y^2 + 11xz + 7yz + 6z^2 + 14x + 8y + 14z + 8$. Such eliminations lead to

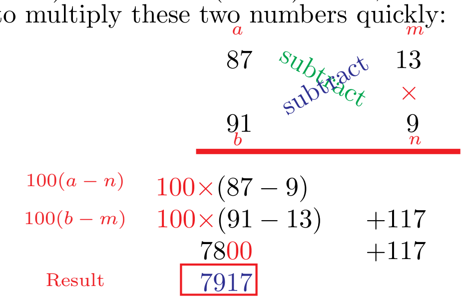
$$Q(x,0,0) = 3x^2 + 14x + 8 = (x + 4)(3x + 2)$$
$$Q(0,y,0) = 2y^2 + 8y + 8 = (2y + 4)(y + 2)$$
$$Q(0,0,z) = 6z^2 + 14z + 8 = (3z + 4)(2z + 2).$$

Using a completion method similar to Example 1, we obtain

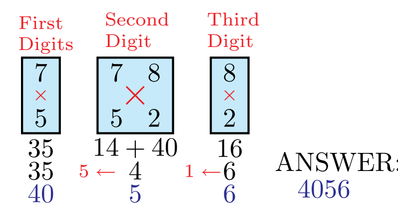
$$Q(x,y,z) = (x + 2y + 3z + 4)(3x + y + 2z + 2).$$

It is easy to verify that this is indeed a factorization of the polynomial $Q(x,y,z)$.

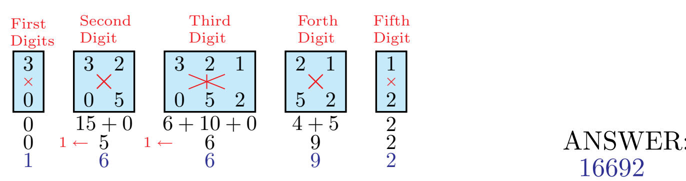
Example 3. In conventional arithmetic, there is no shortcut to multiplying the number $a = 87$ by $b = 91$; this can be done only by 'long multiplication.' But the Vedic method sees these numbers are close to 100 (i.e., the numbers $m = 100 - a$ and $n = 100 - b$ are relatively small). Since $a \cdot b = a(100 - n) = 100(a - n) + mn = 100(b - m) + mn$, there is a very simple way to multiply these two numbers quickly:



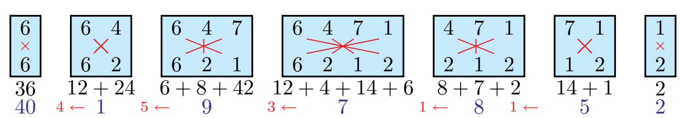
Example 4. Another way of doing a similar multiplication is illustrated below, where we show how to compute the product 78×52 using vertical and crosswise multiplication:



This method can be used to multiply large numbers as well. Let us, for instance, compute the product 321×52:



The product 6471×6212 can be computed in a similar way:



We obtain the answer 40197852.

There are several books written about this fascinating subject, including Vedic Mathematics, by Jagadguru Swami Sri Bharati Krsna Tirthaji Maharaja. Also see the Internet web sites `http://www.vedicmaths.org` and `http://www.mlbd.com`. In many schools, the Vedic system is now being taught to students. "The Cosmic Calculator," a course based on Vedic math, is part of the National Curriculum for England and Wales.

Solar Eclipses for Beginners

by
Ari Stark†

If the article on eclipses in our last issue was too "technical" for you, maybe this simpler one will help to ease you into that fascinating subject. To get oriented, let us look at the following picture, which shows the sun (yellow), the moon (gray), but no Earth. Instead, it has three little square dots (red, orange, blue) in the gray areas behind the moon—they represent possible locations for an observer.

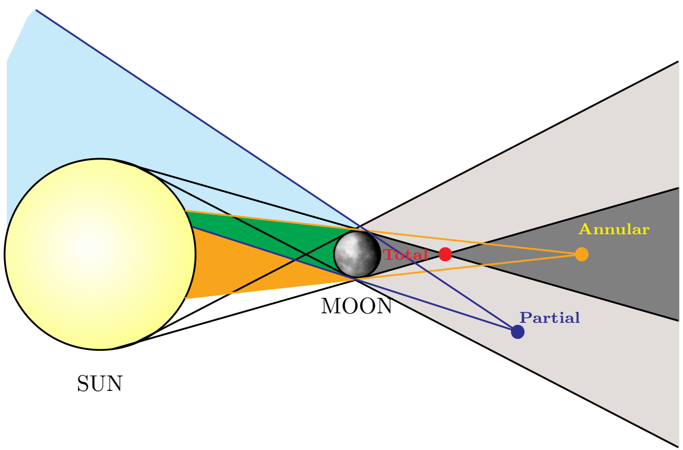


Figure 1

Three types of solar eclipses: total, annular, and partial.

An observer at the red dot would have his view of the sun just barely blocked by the moon. If he displaced himself toward the moon, this *total* eclipse would only become more so; if he went the other way (cf. orange dot) the moon would no longer cover the whole sun, but its shadow would still appear as a complete black disk on the sun (*annular* eclipse). An observer in the light gray area (cf. blue dot) would see only a part of the moon's shadow taking a bite out of the upper or lower half of the sun (*partial* eclipse).

Since all three types occur on Earth, the distance moon-Earth must be somewhat variable—but let us imagine a world without annular eclipses: both sun and moon orbit the Earth in concentric circles at uniform speeds, watched by an observer at the immobile center of the Earth, suspended so as not to notice the daily spin. As we shall see in retrospect, this simple model yields results consistent with those obtained in the last issue. For starters, let's try to understand what's going on in principle.

If the moon were orbiting the Earth in the same plane as the sun, it would get in the sun's way once every month (when it is "new"), and throw its shadow on our central observer. In fact, the latter has long ago charted the sun's course (the "ecliptic") on the inside of the hollow sphere that houses him, and measured the moon's farthest separation from this course to be 5°. Therefore he concludes that the moon's orbital plane is tilted about 5° with respect to the sun's. The intersection of those two planes is called the nodal line, and eclipses occur if sun and moon simultaneously get close to it, i.e., the angle $\beta = \angle SEM$ (sun-earth-moon) is small. Here is the picture.

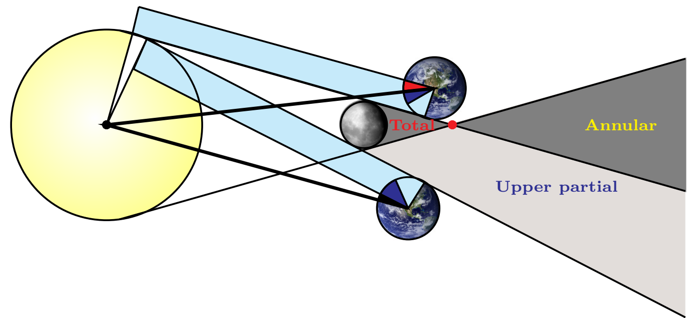


Figure 2

Earth at two different points of the moon's shadow cone.

Wild, eh? That's because the Earth is shown twice, once as it touches the region of total eclipses, and another time for partial eclipses. In the second case, β is easier to study: it is a certain "blue angle" γ (to be identified presently) *plus* a little black sliver ε whose tangent is $(R - R_0)/D$, where R is the radius of the sun, R_0 that of the Earth, and D the distance sun–Earth. The blue angle γ is defined by the Earth–moon axis and the crossing tangent line; it is shifted into the light blue rectangular strips (look at the lower one!) by parallelity. The reason it does not look entirely blue in the upper strip is that a certain red sliver δ, of tangent $(R + R_0)/D$, has been *subtracted* from it.

Since the apparent sizes of sun and moon, as seen from Earth, are very nearly equal (watch a total eclipse!), we have $R/D = r/d$. We really don't need D itself, but the ratio $R_0/D = o$. In these terms the values of β in the two cases are therefore $\gamma + \varepsilon$ and $\gamma - \delta$, respectively, where $\tan \varepsilon = r/d - o$ and $\tan \delta = r/d + o$.

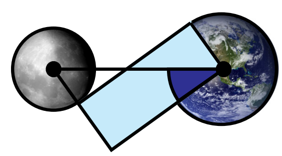


Figure 3

The blue angle γ

As Figure 3 shows, the blue angle γ itself sits in a right triangle opposite a side of length $R_0 + r$, where r is the radius of the moon. Therefore $\sin \gamma = (R_0 + r)/d$, where d denotes the distance between Earth and moon.

In summary: β must be smaller than $\gamma + \varepsilon$ for a partial eclipse, and smaller than $\gamma - \delta$ for a total or annular eclipse, where γ, δ and ε are as defined above.

So far nothing but description—now to the numbers. The following distances are measured in moon radii (mr):

- $d = 220$ mr: distance from earth to moon,
- $R_0 = 3.67$ mr: radius of the earth.

These approximate values are good enough for us—and obtainable by fairly simple observations. You can estimate d by watching total solar eclipses: if you time them, you can see the moon's disk crossing the entire sun in about an hour. This means that the moon is cruising along its path in the sky at the rate of about one diameter an hour. It take four weeks for the moon to do one full turn around Earth: that makes 28×24 hours—and hence the same number of moon

† **Ari Stark** is a pen-name here used by Klaus Hoechsmann (cf. page 24) in honour of Aristarchos, the pioneer of such considerations. His E-mail address is hoek@pims.math.ca.

diameters! To estimate the radius of the circuit, we divide by six and thus obtain $28 \times 4 = 112$ moon diameters (= 224 mr) for the distance Earth-moon. Not quite: a full circuit of the moon as seen from space is only 27 days and eight hours. You can check that by watching the moon wander through the Zodiac—whence the name sidereal month (Latin *sidus* = constellation). Our lunar month (one new moon to the next) is longer because we have to wait for the moon to catch up with the moving sun.

From the moon's cruising speed, Earth's radius can be found by timing total *lunar* eclipses. Here the moon takes about three hours and 40 minutes to tunnel through Earth's shadow—at a speed of one diameter per hour. If, as they tell us, the sun is "much farther" away than the moon, the earth's shadow will be roughly as wide as the earth itself, i.e., approximately 3.67 moon diameters. That is why we put $R_0 = 3.67$ mr.

The only important fact about D is that it is so much bigger than R_0 as to make $o = R_0/D$ negligible in comparison with r/d. Putting $D/d = K$, it is not hard to see that o will be $(367/K)\%$ of r/d. Modern astromers assure us that K is about 400, so all is well. But with their simple instruments, the ancient Greeks obtained values ranging from 20 (Aristarchos) to 1000 (Eratosthenes). That K is "big" is clear to anyone who has observed a half moon high in the sky with its bright half pointing westward almost horizontally instead of pointing at its light source, the setting sun.

Using the values obtained so far, we get $\sin\gamma = 4.67/220$; whence $\gamma = 1.2°$. To estimate the slivers to be added and subtracted from it, we simply omit o and stay with $r/d = 1/220$, which is the tangent of 0.26°; in other words, we estimate $\varepsilon = \delta = 0.26°$. Hence β must be smaller than 1.46° for a partial eclipse, and smaller than 0.94° for a total one. To ensure that a quick brush with the shadow of the moon is not counted as a partial eclipse, let us adjust the 1.46° downward, say, to 1.4°. Question: How often—in all the tuneful turning of the sun—does β fall below these bounds?

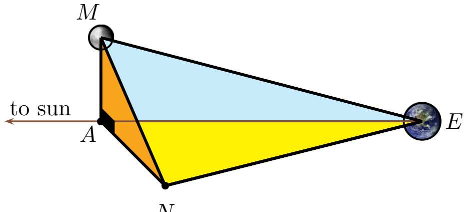


Figure 4

Tetrahedron $NAME$.

To answer it, we shall relate β to an angle α in the orbital plane of the moon. Consider the tetrahedron $NAME$ shown on Figure 4, where the orbital planes of sun and moon are given by the triangles, ANE and NEM, respectively, and NE is the nodal line; A is the "apparent sun," a point on the line SE that lies in the plane perpendicular to NE. Of course, $\beta = \angle AEM$, $d = ME$, and we define α to be $\angle MEN$. We shall say: "the moon is overtaking the sun," when M is directly above A, i.e., when $\angle MAN = 90°$. At that point, all faces of $NAME$ are right triangles. In particular, $MA = d\sin\beta$ and $MN = d\sin\alpha$. On the other hand, $MA = MN\sin 5°$, because $\angle MNA$ is the famous 5° separation between the orbital planes. Putting it all together, we get

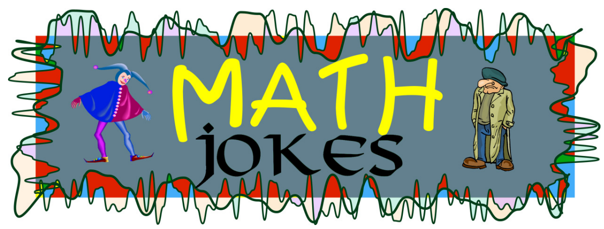
$$\sin\beta = \sin\alpha \cdot \sin 5°.$$

Since $\sin 5° = 0.087$, it follows that β is less than 0.94° or 1.4°, respectively, if α is less than 10.5° or 16°, respectively. We can now gauge the likelihood of an eclipse by angles in the moon's orbital plane. Thus, we have a "partial danger zone" of 16° and a "total danger zone" of 10.5° on either side of the nodal line: if the moon overtakes the sun inside these zones, there is an eclipse somewhere on Earth. When projected onto the ecliptic, these zones practically retain their size since $\cos 5° = 0.996$ is so close to 1.

As the sidereal month of 27.3 days is well within the interval of 32 days the sun must spend in that partial danger zone, one partial eclipse is certain in that time, even two of them are possible—and this situation repeats half a year later, when the sun gets to the other end of the nodal line.

What is the probability of an eclipse on a day chosen at random? Well, the sun must be in its danger zone of 64° out of 360°, and the moon must overtake it on that particular day out of the 27.3 days per circuit. Probability: 8 out of 45 times 27.3—or 1/154. What about total eclipses? The same thing multiplied by 42/64—or 1/234.

This is as far as simple observations will take us. Not bad—but the article by Hermann Koenig in our last issue contains much more information: let's go and look at it again.



The shortest of all math jokes: Let $\epsilon < 0$.

A newlywed husband was discouraged by his wife's obsessions with math. Afraid of playing second fiddle to her profession, he finally confronted her:

"Do you love math more than me?"

"Of course not dear, I love you much more."

Happy, although skeptical, he challenged her: "Well then, prove it."

Thinking a bit, she responds, "Okay, let epsilon be greater than zero. . . "

Math problem? Call 1-800-$[(10x)(13i)^2]$-$[\sin(xy)/2.362x]$.

Applications and Limitations of the Verhulst Model for Populations

Thomas Hillen†

In this article, I use the ongoing discussion about mathematical modelling of historical data as an opportunity to present a classical population model—the *Verhulst* model for self-limited population growth (Verhulst 1836 [3]). I will introduce scaling techniques and demonstrate the method of *perturbation expansions* to understand the usefulness and the limitations of this model.

We assume that $u(t)$ describes the size of a population at time t. The Verhulst model (or *logistic growth* model) is a differential equation, which relates the change in population size over time, du/dt, to birth and death events that occur over time:

$$\frac{d}{dt}u(t) = ru(t)\left(1 - \frac{u(t)}{K}\right), \tag{1}$$

where r is the per capita *birth rate* and K is the *carrying capacity*. The parameter K is a measure of the available resources. If a population reaches the size K, then all resources are used to keep the population level at K and no further growth is possible. If we use this model to describe the development of a population, which at initial time $t = 0$ has the size M, then the solution is given by:

$$u(t) = \frac{KM}{(K-M)e^{-rt} + M}. \tag{2}$$

Exercise 1: *Check that $u(t)$ is indeed a solution of the Verhulst model. What does $u(t)$ look like? What happens for $t \to 0$ and what happens for $t \to \infty$? Use a computer to graph this function. Play around with the parameters r, K, and M. What do you observe?*

The Verhulst model can be used, for example, to describe experimental data collected by Gause [2] on the growth of bacteria populations *Paramecium aurelia* and *Paramecium caudatum*. The time unit is days, and the populations are measured in individuals per cm^3. In the following graph, you see the data for these two measurements and the solution curves of the corresponding model. For *P. aurelia*, we have a birth rate of $r = 0.79$ per individual per day and a carrying capacity of $K = 543.1$ individuals per cm^3. For *P. caudatum*, we have $r = 0.66$ and $K = 202.6$.

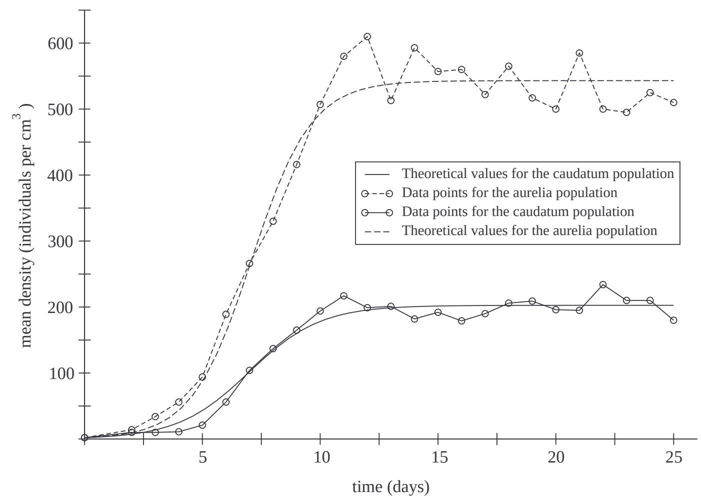


Figure 1

Comparison of the Verhulst model with Gause's experimental data for the growth of *P. caudatum* and *P. aurelia* (from [1]).

For both experiments, we observe nearly exponential growth at the onset, which eventually goes into saturation and converges to its carrying capacity.

The Verhulst model is a *deterministic* model, which means it does not include any stochastic components. Hence the fluctuations about the level of K, as seen in the measurements, cannot be explained by this model.

Let's now have a more theoretical glance at the general Verhulst model (1) and study two special cases: Case 1: *unlimited resources* and Case 2: *small resources and small birth rate*. I will use these cases to illustrate the method of *perturbation analysis*, which is widely used in applied mathematics.

Case 1: Unlimited resources. We assume that the carrying capacity K is large compared to typical population sizes. Then $\varepsilon = 1/K$ is a small quantity. We rewrite the Verhulst model as

$$\frac{d}{dt}u = ru - \varepsilon r u^2. \tag{3}$$

Now we consider a *perturbation expansion in ε:*

$$u(t) = \sum_{j=0}^{N} \varepsilon^j\, u_j(t), \qquad N \geq 2 \tag{4}$$

and we try to determine the coefficient functions $u_j(t)$ for $j = 1, \ldots, N$. It might look funny to replace one function $u(t)$ by a whole set of unknown functions $u_j(t), j = 1, \ldots, N$. In the expansion above, terms are arranged according to their relative importance. Since ε is a small number (e.g., $\varepsilon = 10^{-2}$), the higher exponents, ε^j, are even smaller. Hence we expect that the main information is carried by the first term $u_0(t)$. The other terms are corrections to u_0. We call $u_0(t)$ the *leading-order term* and $u_j(t)$ for $j \geq 1$ the *j-th order correction.*

Now we use the above expansion (4) and plug it into the equation (3)

$$\sum_{j=0}^{N} \varepsilon^j \frac{d}{dt} u_j = \sum_{j=0}^{N} r\varepsilon^j u_j - \varepsilon r \left(\sum_{j=0}^{N} \varepsilon^j u_j\right)^2.$$

† **Thomas Hillen** is a professor in the Department of Mathematical Sciences at the University of Alberta.
His web site is http://www.math.ualberta.ca/~thillen/ and his E-mail addresss is thillen@math.ualberta.ca.

To find the coefficient functions u_j we compare orders of ε, which means we collect terms that have the same power of ε^j:

$$\begin{aligned}
\varepsilon^0: &\qquad \frac{d}{dt}u_0 = ru_0,\\
\varepsilon^1: &\qquad \frac{d}{dt}u_1 = ru_1 - ru_0^2,\\
\varepsilon^2: &\qquad \frac{d}{dt}u_2 = ru_2 - 2r(u_0u_1),\\
\varepsilon^3: &\qquad \frac{d}{dt}u_3 = ru_3 - r(2u_0u_2 + u_1^2),\\
\dots &
\end{aligned}$$

These are differential equations for the coefficient functions u_j. We need to specify some initial conditions. The leading-order term is supposed to carry the major information, hence it is reasonable to assume that initially

$$u_0(0) = M, \qquad u_j(0) = 0, \quad \text{for} \quad j = 1, \dots, N.$$

We observe that the linear growth model

$$\frac{d}{dt}u_0 = ru_0$$

appears as a leading-order approximation (ε^0-equation) to the Verhulst model. It is solved by

$$u_0(t) = Me^{rt}.$$

Now we can use this function $u_0(t)$ to solve the ε^1-equation for the first-order correction,

$$u_1(t) = M^2(e^{rt} - e^{2rt}).$$

From there, we can find the higher-order corrections $u_2(t), u_3(t)$, etc. Solving the ε^j-equations in a row we obtain a sequence of approximations to $u(t)$. The *leading-order approximation* is $u_0(t)$, the *first-order approximation* is $u_0(t) + \frac{1}{K}u_1(t)$, the *second-order approximation* is $u_0(t) + \frac{1}{K}u_1(t) + \frac{1}{K^2}u_2(t)$, and so on. Note that we used $\varepsilon = 1/K$ here.

Exercise 2: *Choose a large value for K (e.g., $K = 100$) and use a computer to compare the solution $u(t)$ given in (2) with its leading-order and first-order approximations.*

Case 2: Here we assume that the birthrate r is equal to a small number ε, and also the carrying capacity is small, like $K = \varepsilon/k$ for some constant $k > 0$. Then the ratio $r/K = k$. We use this scaling in Verhulst's model (1) to get

$$\frac{d}{dt}u = \varepsilon u - ku^2.$$

Again we study a perturbation expansion in ε:

$$u(t) = \sum_{j=0}^{N} \varepsilon^j\, u_j(t), \qquad N \geq 2,$$

and we compare coefficients of ε^j.

$$\begin{aligned}
\varepsilon^0: &\qquad \frac{d}{dt}u_0 = -ku_0^2,\\
\varepsilon^1: &\qquad \frac{d}{dt}u_1 = u_0 - 2ku_0u_1,\\
\dots &
\end{aligned}$$

We use the same initial conditions as in Case 1 above. The equation for the leading-order term u_0 is solved by

$$u_0(t) = \frac{M}{1 + Mkt}.$$

Hence $u_0(t)$ decreases until it reaches 0 as $t \to \infty$.

In these two cases, we see that one model—the Verhulst model—can predict complimentary behavior. Depending on the relative size of the parameters, we obtain, in leading order, exponential growth in Case 1 and decay to 0 in Case 2.

Exercise 3: *Try another scaling.*
(a) What happens to leading order if r is small but K is large?
(b) Use this method to consider time scaling. For example, let's define a "slow" time scale $\tau = \varepsilon t$ and then study $U(\tau) = u(\tau/\varepsilon)$. By using the chain rule, you can derive a differential equation for $U(\tau)$. Then study a perturbation expansion.
(c) What happens with a "fast" time scale like $\theta = t/\varepsilon$?
(d) Try other scaling then interpret your results.

References

[1] G. de Vries, T. Hillen, M. Lewis, M. Li, J. Müller, and B. Schönfisch. A Short Course in Mathematical and Computational Biology. SIAM Publishing (2003). In print.

[2] G.F. Gause. The Struggle for Existence. Hafner (1969).

[3] P.F. Verhulst. Notice sur la loi que la population suit dans son accroissement. Corr. Mat. et Phys. 10 (1838), pp. 113–121.

APPENDIX TO THE ARTICLE

HOW TO READ A MATHEMATICAL PAPER
by Thomas Hillen

Once I asked a student how she approached reading a mathematical paper and she said: "I sit at my desk and stare at it for a very long time... eventually I will understand... hopefully."

Well, there are certainly many ways to read a mathematical paper. The following method works pretty well for me:

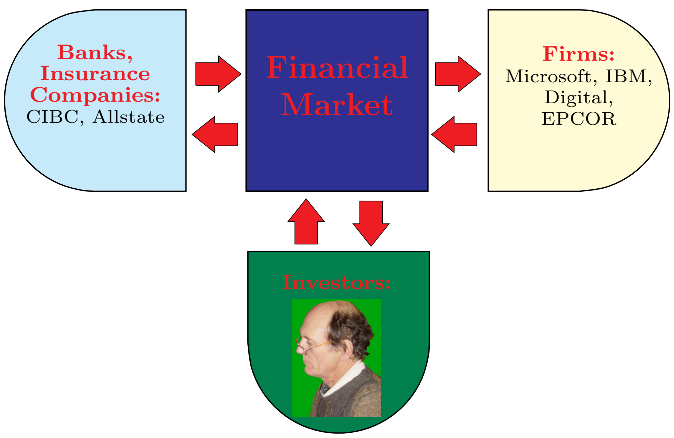
1. Read the paper straight through. Don't bother about the mathematical details. Try to understand what it is about, what are the results, what is the point?
2. Now check the details. Take some blank paper and a pencil and follow all the calculations and modifications. This is the only way to gain a deep understanding of the paper!
3. After checking all the details, read it again. What methods are used? What is the basic idea behind the proof(s)?
4. If you wish, go further. Ask: can it be generalized? Can the method or result be applied to some other problem? Can I shorten the proof? Would a different method be more (or less) efficient? Ultimately, you will start your own research...

You can use the above paper on Verhulst's model to test this method. You don't need to attempt the exercises on first reading; that can wait until step 2. The exercises are essential to understand the article. Have fun!

Mathematics in Today's Financial Markets

Alexander Melnikov†

Market—this idea is usually associated with institutions, people, and actions involved in trading valuables. The valuables, or assets, are called securities (soon we shall talk about these in more detail). The place where we trade them is called a financial market. Not only people, but banks, firms, investment and insurance companies, pension funds, and other structures participate in financial markets.



Buying and selling, owning and loaning assets, receiving dividends, and consuming capital are some of the activities that take place at financial markets. In modern days, these activities require serious quantitative calculations, which we cannot conduct unless we "idealize" the market. For instance, we must assume that all operations and transactions take place immediately (this is called a liquid market) and that they are free (the notion of a frictionless market). Securities are the basis for a financial market. Securities come in many shapes and kinds, but the main ones are stocks and bonds.

Stocks are securities that hold a share of the value of the company (the words stocks and shares are used interchangeably). A company issues stock when it needs to raise capital (money). People buy stock and thus own a "piece" of a company. This ownership gives stockholders the right to make decisions in the way the company is governed and to receive dividends based on the amount of stock a person has.

Bonds are other instruments that a government or a company issues when it needs to raise money. In effect, the buyers of bonds lend their money to the institutions that issue bonds. But such debt must be paid off, and this is done in two ways. Unlike stocks, bonds have an expiry date, indicating when the original borrowed amount (nominal value or principal value) must be paid to the lender. In addition, throughout the term of the bond, the lender receives coupon payments according to the "yield" indicated on the bond. The bond yield is a very significant quantitative indicator for financial calculations; it is similar to a bank's rate of interest—the "reward" for investing money in that bank. The bond without coupon payments can viewed as the "money" in the bank account.

Let's say that a bank persuades you to invest your funds (and now, at time 0, you have the amount B_0) in one of its accounts for a certain period of time (one month, three months, one year, etc.) by promising that at the end of this period (time 1), you will receive a risk-free yield, that is, your initial investment will increase by an amount denoted $\triangle B_1$. Note that $\triangle B_1 = B_1 - B_0$, and also $\triangle B_1 = rB_0$, where r is the interest coefficient, or the bank's interest rate.

Depending on whether you decide to reinvest (monthly, quarterly, yearly), you will receive only the initial investment, or that plus the interest you have earned after $n = 1, 2, 3, \ldots$ periods of time (see Figure 1).

$$\begin{cases} B_n = B_{n-1} + rB_0 = B_0(1 + rn) \\ \quad\text{or} \\ B_n = B_{n-1}(1 + r) = B_0(1 + r)^n. \end{cases} \tag{1}$$

The relationship $\frac{\triangle B_n}{B_{n-1}} = \frac{(B_n - B_{n-1})}{B_{n-1}} = r$ characterizes the yield of your investment.

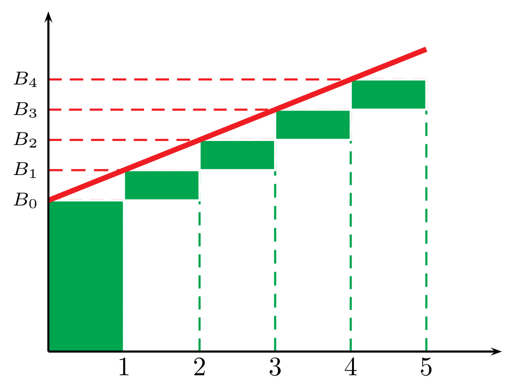


Figure 1
Simple interest—linear growth.

Usually, the rate of interest, or the "yield" on the investment, $r \cdot 100\%$, is stated for a year. We can divide this time period into m smaller periods and calculate the yield (monthly, quarterly, semi-annually, etc.) at the end of each period, according to the stated annual rate. More frequent compounding leads to an increase in the investor's capital; the amount $B_n^{(m)}$ is given by:

$$B_n^{(m)} = B_0 \left(1 + \frac{r}{m}\right)^{mn}. \tag{2}$$

If we subdivide the year into more and more periods, so that m approaches infinity, then $B_n^{(m)}$ approaches $B_0 e^n$. In

† **Alexander Melnikov** is a professor in the Department of Mathematical and Statistical Sciences at the University of Alberta. His web site is `http://www.math.ualberta.ca/Melnikov_A.html` and his E-mail address is `melnikov@ualberta.ca`.

other words, the "limiting" amount of money in the bank account is

$$\lim_{m\to\infty} B_n^{(m)} = B_0 e^n. \tag{3}$$

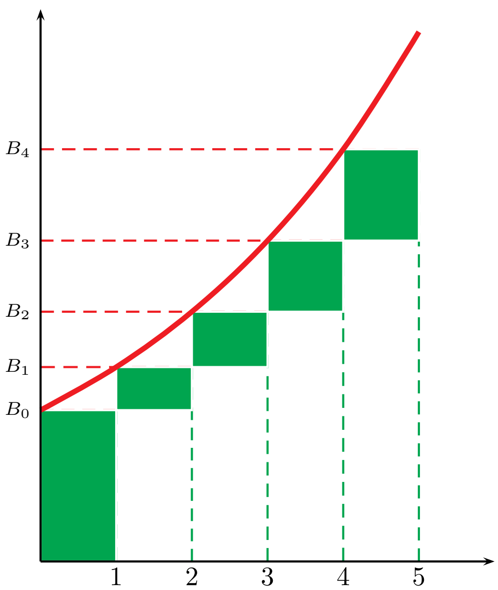


Figure 2
Compound interest—non-linear (exponential) growth.

This implies that the relative yield of such an investment is constant and equals the interest rate r.

The three methods of calculating interest discussed above are called simple, compound, and continuous. Formulas (1), (2) and (3) provide ways to calculate the amount in the investor's bank account and clearly show the dependence of the value of money on time.

On the other hand, we have the bank's interest rate R, so that if we invest the amount $(1-R)B_1$, at time 1, say, after one year, we will receive the amount B_1. This is equivalent to the issuance of a bond with a nominal value B_1 (to be paid to the bond holder at the end of this year), but *now* the bond sells for a lower price (decreased by the amount of the lending rate R over one period, that is, $m=1$). So today's price is determined by the formula $(1-R)B_1$, which is equal to the discounted price $B_1/(1+r)$. Therefore, we can view the bank account as a coupon-free bond in the sense of a risk-free asset of the financial market. The lack of, or very small, changes in interest rates characterize the stability of financial and economic systems, for which the corresponding bank account serves as the basic non-risky asset. Reality shows that such suggestions present limits in the idealization of mathematical models for financial markets.

Formulas (1), (2) and (3) show time evolution of the value of money, presenting difficulties in the calculations of annuities. These are periodic payments to be made in the future (such as rent), denoted by $f_0, f_1, f_2 \ldots, f_n$, whose values we need to know today. According to the compound interest formula, we calculate the value of the k^{th} payment as $f_k/(1+r)^k$. Thus, the cost of all future payments today is given by the sum

$$f_0 + \frac{f_1}{(1+r)^1} + \frac{f_2}{(1+r)^2} + \ldots \frac{f_n}{(1+r)^n}.$$

These and similar arithmetic calculations determining rent payments were the only functions of mathematics in finance until the middle of the 20th century.

After the risk-free bank account, the second basic element of a financial market is a stock, which is much more volatile and thus is called a risky asset. Let S_n denote the price of the stock at time n. We determine the yield of a stock during any time period by $\rho_n = (S_n - S_{n-1})/S_{n-1}$, where $n = 1, 2, \ldots$. Then stock prices satisfy this equation:

$$S_n = S_{n-1}\left(1+\rho_n\right). \tag{4}$$

Bank account balance (1), interest rate r and stock price (4), for changing yield ρ_n form the mathematical model of a financial market.

Many factors, often very difficult to determine, cause changes in stock prices S_n. We refer to these factors as randomness and call S_n (and thus ρ_n) random variables. Just like the yield of a bank account, $r = \triangle B_n/B_{n-1}$, ρ_n is the changing yield of a risky asset (stock in our example). Note that since ρ_n changes every time period (at each $n = 1, 2,$ etc.), we can take all these different values of ρ and calculate their mean μ, and individual yield values will lie below and above the mean. As we shorten our time periods (for instance, instead of observing changes in stock prices and thus monthly or weekly yields, we record the changes hourly or even every minute), we see that the up-and-down movements of the stock's yield become more and more chaotic. The picture below shows a possibility of such limiting behaviour of yield-per-time, where discrete time periods, divided again and again, become a continuous timeline.

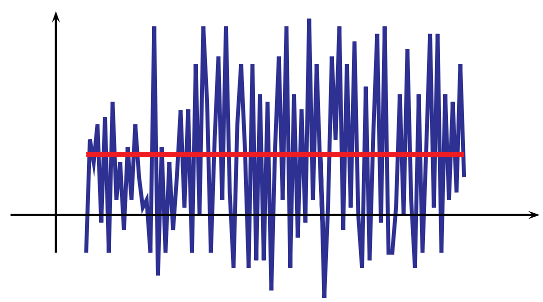


Figure 3
Varying yield values and their time mean.

Formally, in the model with continuous time, at any moment in time t, the limiting yield equals the sum

$$\mu + \sigma\widetilde{W}_t, \tag{5}$$

where μ is the mean yield, σ volatility, and $\widetilde{W}_t$ represents Gaussian "white noise," a notion used in math and physics to describe chaotic, irregular movements.

The pairs of formulas (1) and (4), and (3) and (5), respectively constitute the binomial and diffusion models of the financial market and frequently are called the Cox–Ross–Rubinstein model and the Black–Scholes model.

Further, a participant in the securities market has to invest his or her resources into assets available in this market, choosing certain quantities of different assets. We refer to this process as forming one's investment portfolio. Deciding how much of which assets to include in the investment portfolio is the essence of managing capital. Any changes to the contents of the portfolio should limit or minimize the risk from financial operations; this is called hedging the portfolio.

Among all investment strategies, we separate those that bring profit without any initial expense. The possibility of such strategies reveals the presence of arbitrage in a financial market, which means that the market is unstable. The models we discussed above are idealized, in that they do not allow any arbitrage opportunities.

Developments in the financial market now give its participants access to instruments evolved from basic stocks and bonds. Forwards, futures, and options, called derivative securities, attract investors with lower prices. Derivative securities increase the liquidity of the market and function as insurance against losses from unsuccessful investments.

For example, consider company A, which wishes to buy stock of company B at the end of this year. The price of B's stock can either increase or decrease. So, to insure itself against higher prices, company A signs a forward contract with company B. According to the contract, A will buy B's stock at a predetermined and fixed price F at the end of the year.

Now consider another case. Say company A already has B's stock, so A wisely wants to insure itself against the losses it would incur if the price of B's stock falls. Therefore, A purchases a seller's option from B. This agreement grants A the right to sell B's stock at a predetermined and fixed price K at the end of the year. For the opportunity to do so, A pays B a price for the contract—a premium.

A future contract is similar to a forward contract, but rather than being written by the two participating sides directly, it is made through an exchange—a special organization for managing the trade of various goods, financial instruments, services, etc. At an exchange, all commercial operations are done by brokers, or intermediaries, who bring together individuals and firms to make contracts.

The first exchange specializing in the trade of options, CBOE (Chicago Board Option Exchange), opened on April 26, 1973, and by the end of the first day of work as many as 911 contracts were signed (one contract equals 100 shares). Since then, the derivative securities markets have grown fast. The huge capital of more and more participating firms and the astonishing volume of contracts being signed increases the volatility of derivative securities markets, thus increasing the randomness factor in the determination of prices of traded assets. Therefore, appropriate stochastic models have become necessary for the valuation of assets. Today, probability theory and mathematical statistics are used to develop such models for financial markets.

In the entire spectrum of securities, the most significant one, mathematically, is an option—a derivative security that gives the right to buy stock (this is a "call" option) at a predetermined price K at the termination time T. (Note that the right to sell stock is a "put" option). The exercise of a call option demands payment of $(S_T - K)^+$, which is the greater of $(S_T - K)$ and zero. Likewise, we have $(K - S_T)^+$ with a put option. The main problem, both practically and theoretically, is this: what should be the *current* price P_T of the contract C_T? We only need to find C_T or P_T since $P_T = C_T - S_0 + \frac{K}{(1+r)^T}$ or $P_T = C_T - S_0 + \frac{K}{e^{rT}}$.

In the case of the binomial model, we have two possibilities for the stock price at the end of a period: either it will go up with the probability p, or it will go down with the probability $1-p$. So the stock yield ρ will take on values b in $p \cdot 100\%$ of cases and a in $(1-p) \cdot 100\%$ of cases, with $b > r > a > -1$. The exact answer for the price of the call option in the binomial model is given by the famous formula of Cox–Ross–Rubinstein (1976):

$$C_T = S_0 \sum_{k=k_0}^{T} \frac{k!}{T!(T-k)!} \widetilde{p}^k (1-\widetilde{p})^{T-k}$$
$$-K(1+r)^{-T} \sum_{k=k_0}^{T} \frac{k!}{T!(T-k)!} (p^*)^k (1-p^*)^{T-k},$$

where $k!$ is the product $1 \cdot 2 \cdot \ldots \cdot k$, $p^* = \frac{(r-a)}{(b-a)}$, $\widetilde{p} = p^* \frac{(1+b)}{(1+r)}$, and k_0 is the smallest integer j that makes the quantity $S_0(1+a)^T \left(\frac{1+b}{1+a}\right)^j$ greater than K.

But initially, the answer was found for the diffusion model by Black, Scholes, and Merton in 1973:

$$C_T = S_0 \Phi \left(\frac{\ln \frac{S_0}{K} + T(r + \frac{\sigma^2}{2})}{\sigma\sqrt{T}} \right)$$
$$-K e^{rT} \Phi \left(\frac{\ln \frac{S_0}{K} + T(r - \frac{\sigma^2}{2})}{\sigma\sqrt{T}} \right),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\, dy$$

is the error function corresponding to the standard normal distribution. The significance of this discovery was acknowledged with the Nobel Prize in economics in 1997.

Let us remark that, historically, the first strictly mathematical work in calculating options, the *Theory of Speculations*, was written by L. Bachelier in 1900. However, no one saw the significance of these calculations at the time, and it was only in the middle of the 1960s famous economist Samuelson "rediscovered" Bachelier's paper and introduced the more natural market model (5), known today as the formula of Black, Scholes, and Merton.

As we have mentioned already, options and other derivative securities can function as insurance. Unlike traditional insurance, where a client "sells" his risk to some insurance company, insurance through options (hedging) allows putting this risk in the financial market with the opportunity to watch stock prices and adequately react to changes in the market situation. In this way, finance and insurance are merged.

Therefore, in a financial market, the risk inherent in any investment portfolio can be managed with the insurance method described above. Insurance derivative securities (insurance forwards, futures, options), have become some of the most popular assets to be traded in the past decade. And the quantitative calculations of premiums (contract prices) and risk are done using a mix of methods in financial and actuarial mathematics.

Am I Really Sick?

by Klaus Hoechsmann†

When Nelly came back from her year in Ladorada, she read in The New York Times that tuberculosis was on the rise again, especially in the part of the world she had just visited. Her doctor explained to her that there was not much to worry about, as only 0.01 percent of the inhabitants of that beautiful and hospitable land were affected, but that it was wise to take a simple test, which he had right there in his office. The next day, he phoned her to say that, unfortunately, she had tested positive. "Does that mean I am really sick?" she asked. "It doesn't look good," said the doctor. "The test is 99.9 percent sure."

Nelly

Her brother Nick, who was an engineer with a good feel for numbers, asked her to quote those figures again, and then burst out laughing: only one in 10 000 Ladoradans is infected, and that dumb test makes one mistake in 1000 trials, see?" Nelly shook her head: "No, I only see a guy who giggles at numbers. Tell me what's so funny." Nick gave her that big brother look: "According to your doctor, the test is right 999 times in 1000, on average. So, if you test 10 000 people from over there, you'll get 10 false alarms and one real case—on average." That evening, Nelly took a long walk in Central Park mulling over Nick's argument that her chances of being sick were only one in 11. What a relief!

She had dinner with her friend Cornelia, who was studying to be a nurse. "I can't get over it," said Cornelia. "In class today we just began to study that new TB outbreak in Ladorada as well as the Kinski Test, and here I am having dinner with a real specimen—that's so cool." Nelly recalled seeing the name Kinski on the box in Dr. Dixit's office. "It's a new test," Cornelia rambled on, "99.99 percent accurate. But don't worry, Nelly: even if you test positive, the chances you are really infected are only 50 percent. That's what our instructor said—always 50 percent—and he does research in a big lab." Nelly was aching to run Nick's logic through her mind. "My doctor said its accuracy was only 99.9 percent," she ventured. "So what?" Cornelia shot back, "In either case, we have near certainty. Remember, you always have a 50–50 chance to be a false positive." Nelly remarked that Nick had convinced her that the chance was 10 in 11, if Kinski was only 99.9 percent certain.

Though she had always had an eye on Nelly's brother, Cornelia now snapped that he was just an electrical engineer. "What does he know about epidemiology?" She spoke that word with the solemnity of a neophyte. To save the evening, the two women opened another bottle of wine.

As she was waking up the next morning, Nelly was able to reconstruct Nick's argument for a test that was 99.99 percent accurate. It would make one false diagnosis in 10 000 cases, and in Ladorada there would be one real TB-carrier among them—on average, as Nick would say. Hence you would likely find two "positives" in that crowd. The reasoning behind Cornelia's "50–50" chance was therefore the equality of the infection rate in the country and the failure rate of the test, namely one in 10 000. Good! No, bad!

It meant that she had better do something: 50 percent was too close for comfort. She got an appointment with an X-ray lab, but only after the impending long weekend, most of which she spent practicing her violin. After all, TB was not AIDS, although the new strain was said to be particularly resistant to treatment. But time and again, she was drawn to search in the Internet for news on TB in Ladorada.

On Monday, still a holiday, she found a reputable Spanish site describing the uneven spread of tuberculosis in Ladorada: the capital was stricken 10 times harder than the country as a whole. "One in 1000 for the likes of me," she thought, because she had spent almost all her time in Hermosa. She reviewed Nick's reasoning: if 10 000 inhabitants of that city were tested, 10 true positives would turn up—on average—because of the infection rate, and one false positive because of the margin of error in Kinski. The tables had turned: her chances of being healthy were down to one in 11. What a bummer! Just then Cornelia phoned to say how sorry she was to have been so snarky about Nick. Nelly told her about the new odds.

Cornelia

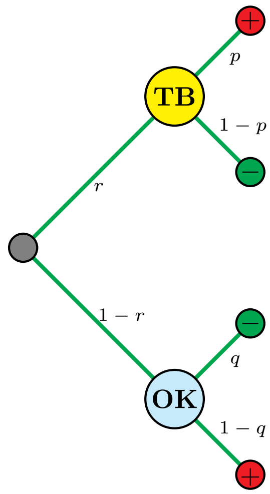
"Don't worry so much," Cornelia suggested. "The experts say the odds are 50–50 for false positives; that's what you should go by, instead of confusing yourself with simplistic calculations." The conversation ended with some chatter about Nick's vulnerability to predatory females. Quite a pair of health professionals, Dr. Dixit and Cornelia, thought Nelly. But the word "simplistic" struck a chord. What about the false negatives—sick people given a clean bill of health by the test. Wouldn't they diminish the 10 "true" positives?

Not by much, of course, but Nelly was suddenly more interested in the calculation than in her own health. She phoned her friend Fatima, a graduate student in statistics, and was glad to find her at home. After hearing what the problem was, Fatima invited her over for tea. "It's a classic," she smiled, "I have to explain it in my tutorials every single year, so I made this slide to put on the overhead." Nelly vaguely made out some letters, lines, and coloured dots. "We'll go over it after tea," said Fatima.

"See this yellow dot here? It represents the infected part of the population, and the letter r stands for the odds of being infected, 1/1000 in your case, but it could be 38/31570 or something crazy like that: it's best to think of it as a percentage. And the $1 - r$ is the opposite percentage, the chance of being uninfected, hence it's connected to the blue dot labelled OK." All that was straightforward, but didn't tell you anything, thought Nelly. "How come the yellow dot

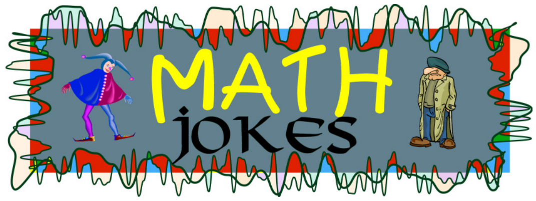
† **Klaus Hoechsmann** is a professor emeritus at the University of British Columbia in Vancouver, B.C. You can find more information about the author and other interesting articles at: http://www.math.ubc.ca/~hoek/Teaching/teaching.html. His E-mail address is hoek@pims.math.ca.

is labelled TB?" she asked. "Oh, just because it has only two letters. The so-called infection could be any hidden condition you want to ferret out with your test—like a secret yearning," Fatima whispered, "Say, for chocolate. But let's get back to the test."



"The test marks you either as positive—those are the red dots—or negative, shown here in green." Now Nelly began to catch on. "And there are two of each since you could be a true or a false positive or negative. Could I try to explain the rest of the diagram?" Fatima was delighted to have such an eager student. "If you are yellow, the chances of being correctly identified are p; if you are blue, they are q. Why are they not the same?" Fatima pointed out that the true yellows were usually easier to identify than the true blues, so p was usually bigger than q."But in the Kinski Test, they are the same: 99.99 percent, aren't they?" said Nelly. "I looked it up after you phoned me," replied Fatima, "p is indeed 99.99 percent but q only 99.9 percent—so the test typically produces one positive out of every 1000 blues." Nelly threw her arms around Fatima: "That means I'm back to 50 percent, if Nick's reasoning is correct. Oh, Fatima, please tell me that it is!" Fatima thoroughly enjoyed being the bringer of glad tidings, but said: "Not quite. Let's work it out all the way."

They tallied the positives: the true ones from yellow were $r \times p$ and the false ones from blue were $(1-r) \times (1-q)$, for a total of $(1-r)(1-q)+rp$. Thus, the chances of a false positive were 1 in $1+rp/(1-p)(1-q) = 1+[r/(1-q)][p/(1-r)]$. "Nick neglected that last factor $p/(1-r)$," said Fatima, "But look at it: 99.99/99.9 in this case. It isn't much of a factor, and that is typical: for any half-decent test it's very close to one." Nelly had tears in her eyes and didn't know whether they came from the 50 percent or from understanding the calculation. "Fatima, you are an angel," she said, "But tomorrow I'll still have myself checked out." Back on the street, the only thing that irked her was that the news would strengthen Cornelia's blind faith.



In a speech to a gathering of mathematics professors from across the United States, a conservative politician warned academics not to misuse their position to force their often extremist political views on young Americans. "It is my understanding," the politician said, "that you frequently teach algebra classes in which your students learn how to solve equations with the help of radicals. I can't say that I approve of that. . . "

It is only two weeks into the term when, in a calculus class, a student raises his hand and asks: "Will we ever need this stuff in real life?" The professor gently smiles at him and says: "Of course not—if your real life will consist of flipping hamburgers at McDonald's!"

Three statisticians go hunting. When they see a rabbit, the first one shoots, missing it on the left. The second one shoots and misses it on the right. The third one shouts: "We hit it!"

An American mathematician returns home from a conference in Moscow on real and complex analysis. The immigration officer at the airport glances at his landing card and says: "So, your trip to Russia was business related. What's the nature of your business?"

"I am a professor of mathematics."

"What kind of mathematics are you doing?"

The professor ponders for a split second, trying to come up with something that would sound specific enough without making the immigration officer suspicious, and replies: "I am an analyst."

The immigration officer nods with approval: "I think it's great that guys like you go to Russia to help those poor ex-commies to get their stock market on its feet. . . "

An investment firm is hiring mathematicians. After the first round of interviews, three hopeful recent graduates—a pure mathematician, an applied mathematician, and a graduate in mathematical finance—are asked what starting salary they are expecting.

The pure mathematician: "Would $30 000 be too much?"

The applied mathematician: "I think $60 000 would be OK."

The mathematical finance person: "What about $300 000?"

The personnel officer is flabbergasted: "Do you know that we have a graduate in pure mathematics who is willing to do the same work for one-tenth of what you are demanding!?"

"Well, I thought $135 000 for me, $135 000 for you—and $30 000 for the pure mathematician who will do the work."

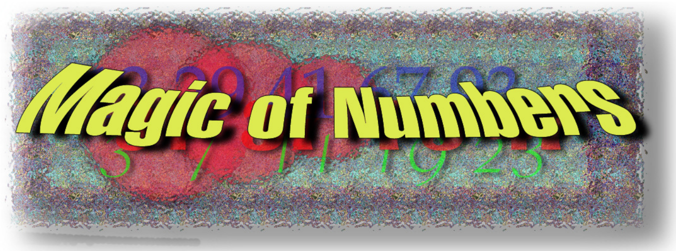
©Copyright 2003
Sidney Harris

Statistics Canada is hiring mathematicians. Three recent graduates are invited for an interview: one has a degree in pure mathematics, one one in applied math, and the third a B.Sc. in statistics. All three are asked the same question: "What is one-third plus two-thirds?"

The pure mathematician: "It's one."

The applied mathematician takes out his pocket calculator, punches in the numbers, and replies: "It's 0.999999999."

The statistician: "What do you want it to be?"



Divisibility by Prime Numbers

Edwin D. Charles† and Jeremy B. Tatum*

1. Introduction

Many of us may remember from our school days being taught how to test whether a given large number is divisible by 2, 3, 5, 6, 8, 9, 10, 11 or 12. Early in 2002, JBT (hereafter "I") wrote to Mr. Charles to ask if he knew of any method to test whether a number is divisible by seven. Not very long afterwards, Mr. Charles replied with a successful test that he had devised. A little while later, he sought to expand upon this; he had almost completed developing a method to test whether a given large number is divisible by a specified prime number when ill-health at the age of 91 obliged him to give up the time needed to complete his work. Mr. Charles died on December 23, 2002. His letters to me on the subject were so clear and organized that they were almost ready for publication as a formal article. With the help of a computer (unavailable to Mr. Charles), I expanded his Table of Moduli (which he had completed by hand up to $p = 29$ and $n = 32$) to $p = 97$ and $n = 100$, made one or two minor modifications, and prepared an article for publication. However, the method described, which constitutes the core of the paper, is that of Mr. Charles, who should therefore be regarded as the principal author.

2. Divisibility by 7, 13, 37, and 73

A rather simple test can be devised to test a large number for divisibility by 7, 13, or 37, and to test for divisibility by 73 is only slightly longer. In this section, we will describe, without explanation, the tests for divisibility by these four numbers. In the next section, we will explain why these tests work and show how to devise a test for divisibility by any prime number. We will supply sufficient data to enable any reader quickly to devise a test for divisibility by any of the prime numbers up to p = 97.

The test number we use in this article will be

$$x = 6986648088495576619729344372307579911.$$

This number is not an arbitrarily-chosen number. We will explain its significance a little later.

To test it for divisibility by 7, 13, or 37, we write the number in groups of three digits:

6 986 648 088 495 576 619 729 344 372 307 579 911.

† **Edwin D. Charles** (1910-2002) was chief electrical engineering draughtsman at the South Eastern Electricity Board in England. He worked on this prime number project at the age of 91.

* **Jeremy B. Tatum** was an astronomy professor at the University of Victoria. During his research career, he discovered several new asteroids. His E-mail address is universe@uvvm.uvic.ca.

Symbolically, let us write this as

$$a_{13}\quad c_{12}b_{12}a_{12}\quad c_{11}b_{11}a_{11}\quad c_{10}b_{10}a_{10}\quad c_9\,b_9\,a_9\quad c_8\,b_8\,a_8\quad c_7\,b_7\,a_7$$
$$c_6\,b_6\,a_6\quad c_5\,b_5\,a_5\quad c_4\,b_4\,a_4\quad c_3\,b_3\,a_3\quad c_2\,b_2\,a_2\quad c_1\,b_1\,a_1.$$

Form the sums:

$$A = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10} + a_{11} + a_{12} + a_{13},$$
$$B = b_1 + b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9 + b_{10} + b_{11} + b_{12},$$
$$C = c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 + c_8 + c_9 + c_{10} + c_{11} + c_{12},$$
$$A' = a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + a_7 - a_8 + a_9 - a_{10} + a_{11} - a_{12} + a_{13},$$
$$B' = b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + b_7 - b_8 + b_9 - b_{10} + b_{11} - b_{12},$$
$$C' = c_1 - c_2 + c_3 - c_4 + c_5 - c_6 + c_7 - c_8 + c_9 - c_{10} + c_{11} - c_{12}.$$

For the number x, these sums have the values

$$A = 80,\quad B = 58,\quad C = 60,$$
$$A' = 0,\quad B' = -20,\quad C' = 2.$$

We now assert:

- x is divisible by 7 iff $r = A' + 3B' + 2C'$ is divisible by 7.
- x is divisible by 13 iff $r = A' - 3B' - 4C'$ is divisible by 13.
- x is divisible by 37 iff $r = A + 10B - 11C$ is divisible by 37.

For our number x, these three expressions have the values -56, 52, and 0 respectively. Since these numbers are, respectively, divisible by 7, 13, and 37, the number x is divisible by 7, 13, and 37. In case one is not sure whether 52 is divisible by 13, one can apply the test again. Thus, for the number 52, $A' - 3B' - 4C'$ is equal to -13, and therefore 52 and x are both divisible by 13.

To test for divisibility by 73, one writes the number to be tested in groups of four:

6 9866 4808 8495 5766 1972 9344 3723 0757 9911

In symbols, ... $d_2c_2b_2a_2\quad d_1c_1b_1a_1$.

We need the sums A', B', C', and $D' = d_1 - d_3 + d_3 - \ldots$ For the number x (as written now in groups of four digits), we find that

$$A' = -14,\quad B' = 10,\quad C' = 12,\quad D' = 12.$$

We now assert:

- x is divisible by 73 iff $r = A' + 10B' + 27C' - 22D'$ is divisible by 73.

In this case, we find that $r = -14 + 100 + 324 - 264 = 146$. If one is not sure whether 146 is divisible by 73, one can check this too: for 146, $r = 6 + 40 + 27 = 73$. Thus 146 and x are both divisible by 73.

3. Divisibility by Any Prime Number

The rationale for the tests we have described is as follows. Let x be an integer of $n + 1$ digits written in the form

$$x = a_n a_{n-1} a_{n-2} \ldots a_2 a_1 a_0,$$

and let p be a prime number. Let

$$r = (10^0 \bmod p)a_0 + (10^1 \bmod p)a_1 + (10^2 \bmod p)a_2 + \dots$$
$$\dots (10^{n-1} \bmod p)a_{n-1} + (10^n \bmod p)a_n.$$

Then, if r is divisible by p, then, and only then, x is divisible by p. To see this, think how one would divide 3256607, say, by 7. One would mutter to oneself: "7 *into* 32 *goes* 4 *and* 4 *over;* 7 *into* 45 *goes* 6 *and* 3 *over;* 7 *into* 36 *goes* 5 *and* 1 *over.*"

For this test to be useful, we need a table of $10^n \bmod p$, and we provide a table for prime numbers from 7 to 97. It is easy to extend this table to as large a value of n as you wish, as will shortly become evident, and it will be found, for example, that $10^{100} \bmod 97 = 9$. This means that if you were to divide the number 10 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 by 97, there would be a remainder of 9. To extend the table, we notice that, as we run our eye down each column, the pattern of entries repeats itself, either with a change in sign or with no change in sign after each period. Indeed, for a given prime number, the period is never more than $p - 1$, and in many cases it is a fraction of this.

p / n	7	11	13	17	19	23	29	31	37	41	43	47
0	1	1	1	1	1	1	1	1	1	1	1	1
1	3	-1	-3	-7	-9	10	10	10	10	10	10	10
2	2	1	-4	-2	5	8	13	7	-11	18	14	6
3	-1	-1	-1	-3	-7	11	14	8	1	16	11	13
4	-3	1	3	4	6	-5	-5	-13	10	-4	-19	-11
5	-2	-1	4	6	3	-4	8	-6	-11	1	-18	-16
6	1	1	1	-8	-8	6	-7	2	1	10	-8	-19
7	3	-1	-3	5	-4	-9	-12	-11	10	18	6	-2
8	2	1	-4	-1	-2	2	-4	14	-11	16	17	-20
9	-1	-1	-1	7	-1	-3	-11	-15	1	-4	-2	-12
10	-3	1	3	2	9	-7	6	5	10	1	-20	21
11	-2	-1	4	3	-5	-1	2	-12	-11	10	15	22
12	1	1	1	-4	7	-10	-9	4	1	18	21	-15
13	3	-1	-3	-6	-6	-8	-3	9	10	16	-5	-9
14	2	1	-4	8	-3	-11	-1	-3	-11	-4	-7	4
15	-1	-1	-1	-5	8	5	-10	1	1	1	16	-7
16	-3	1	3	1	4	4	-13	10	10	10	-12	-23
17	-2	-1	4	-7	2	-6	-14	7	-11	18	9	5
18	1	1	1	-2	1	9	5	8	1	16	4	3
19	3	-1	-3	-3	-9	-2	-8	-13	10	-4	-3	-17
20	2	1	-4	4	5	3	7	-6	-11	1	13	18
21	-1	-1	-1	6	-7	7	12	2	1	10	1	-8
22	-3	1	3	-8	6	1	4	-11	10	18	10	14
23	-2	-1	4	5	3	10	11	14	-11	16	14	-1
24	1	1	1	-1	-8	8	-6	-15	1	-4	11	-10
25	3	-1	-3	7	-4	11	-2	5	10	1	-19	-6

I have indicated in blue the shortest repetition period for each number without regard to changes of sign. The shorter the period, the easier the divisibility test. In order to test our number x for divisibility by 47, for example, we would have

p / n	53	59	61	67	71	73	79	83	89	97
0	1	1	1	1	1	1	1	1	1	1
1	10	10	10	10	10	10	10	10	10	10
2	-6	-18	-22	33	29	27	21	17	11	3
3	-7	-3	24	-5	6	-22	-27	4	21	30
4	-17	29	-4	17	-11	-1	-33	40	32	9
5	-11	-5	21	-31	32	-10	-14	-15	-36	-7
6	-4	9	27	25	-35	-27	18	16	-4	27
7	13	-28	26	-18	5	22	22	-6	-40	-21
8	24	15	16	21	-21	1	-17	23	-44	-16
9	-25	-27	-23	9	3	10	-12	-19	5	34
10	15	25	14	23	30	27	38	-24	-39	-48
11	-9	14	18	29	16	-22	-15	9	-34	5
12	16	22	-3	22	18	-1	8	7	16	-47
13	1	-16	-30	19	-33	-10	1	-13	-18	15
14	10	17	5	-11	25	-27	10	36	-2	-44
15	-6	-7	-11	24	-34	22	21	28	-20	45
16	-7	-11	12	-28	15	1	-27	31	-22	-35
17	-17	8	-2	-12	8	10	-33	-22	-42	38
18	-11	21	-20	14	9	27	-14	29	25	-8
19	-4	-26	-17	6	19	-22	18	41	-17	17
20	13	-24	13	-7	-23	-1	22	-5	8	-24
21	24	-4	8	-3	-17	-10	-17	33	-9	-46
22	-25	19	19	-30	-28	-27	-12	-2	-1	25
23	15	13	7	-32	4	22	38	-20	-10	-41
24	-9	12	9	15	-31	1	-15	-34	-11	-22
25	16	2	29	16	-26	10	8	-8	-21	-26
26	1	20	-15	26	24	27	1	3	-32	31
27	10	23	-28	-8	27	-22	10	30	36	19
28	-6	-6	25	-13	-14	-1	21	-32	4	-4
29	-7	-1	6	4	2	-10	-27	12	40	-40
30	-17	-10	-1	-27	20	-27	-33	37	44	-12
31	-11	18	-10	-2	-13	22	-14	38	-5	-23
32	-4	3	22	-20	12	1	18	-35	39	-36
33	13	-29	-24	1	-22	10	22	-18	34	28
34	24	5	4	10	-7	27	-17	-14	-16	-11
35	-25	-9	-21	33	1	-22	-12	26	18	-13
36	15	28	-27	-5	10	-1	38	11	2	-33
37	-9	-15	-26	17	29	-10	-15	27	20	-39
38	16	27	-16	-31	6	-27	8	21	22	-2
39	1	-25	23	25	-11	22	1	-39	42	-20
40	10	-14	-14	-18	32	1	10	25	-25	-6
41	-6	-22	-18	21	-35	10	21	1	17	37
42	-7	16	3	9	5	27	-27	10	-8	-18
43	-17	-17	30	23	-21	-22	-33	17	9	14
44	-11	7	-5	29	3	-1	-14	4	1	43
45	-4	11	11	22	30	-10	18	40	10	42
46	13	-8	-12	19	16	-27	22	-15	11	32
47	24	-21	2	-11	18	22	-17	16	21	29
48	-25	26	20	24	-33	1	-12	-6	32	-1
49	15	24	17	-28	25	10	38	23	-36	-10
50	-9	4	-13	-12	-34	27	-15	-19	-4	-3

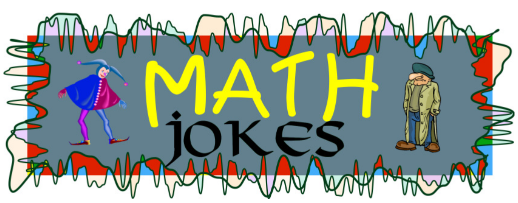
to arrange the number in groups of 23 digits, and form the alternating sums A', B', C',… U', V', W'. We use the alternating (primed) sums because consecutive periods alternate in sign. We then calculate

$$r = A' - 10B' + 6C' + 13D' - 11E' - \cdots + 14W'$$

and deduce that x is divisible by 47 iff r is divisible by 47.

One may ask whether applying the test for divisibility is in practice really faster than actually carrying out the division, particularly since applying the test doesn't tell you what the quotient is. This depends on a number of things, the most important of which is the length of the repetition period. The repetition period for $p = 3$ is nice and short—just one. On the other hand, most of us remember our seven-times table, so dividing by 7 is probably the better bet. The repetition period for $p = 41$ is a bit longer; it is five, so the test is a bit more complicated. But unless you have memorized your 41-times table, the test is probably faster than direct division by 41. What if $p = 97$? The test is a very long one, and it would take some time to work out what the test actually is. Yet can you divide 10^{100} by 97 any faster? We suspect that the only way to find out is to try both methods and time oneself.

One last small detail. The 37-digit number that we chose as a numerical example to illustrate the method is the product of all the prime numbers from 7 to 97 inclusive, and it was worked out using a simple hand calculator carrying ten digits—but how that was done we shall keep as a little secret.



A physics professor is examining three students orally. One of them is in engineering, one is in physics, and the third is in mathematics. The question is the same for each of them: "Which is faster: light or sound?"

The engineering student is first. His answer: "Sound, of course!"

The professor grinds his teeth, but manages to stay calm: "And what makes you think so?"

"Well, whenever I turn on my TV, the screen is still dark when the sound comes…"

"GET OUT!"

The physics student is next, and he answers: "Light, of course!"

The professor is relieved, but nevertheless asks: "And what makes you believe this?"

"That's easy: whenever I turn on my car's sound system, a light goes on before the sound comes…"

"GET OUT!!!"

Before it's the math student's turn, the professor ponders. Maybe his question is too difficult and too abstract. So he gets himself a horn and a flashlight, and when the math student enters his office, he simultaneously blows into the horn and flashes the light at the student.

"Which did you notice first?" he asks. "Light or sound?"

"The light."

"And what is your explanation for this?"

"That's because my eyes are further in front in the head than my ears…"

MATH & MUSIC

This song, adapted from Don McLean's "American Pie" by Lawrence (Larry) M. Lesser from Armstrong Atlantic State University, gives historical highlights of the number π. Visit http://www.real.armstrong.edu/video/excerpt1.html to download a video of Larry performing this and some other math songs. We also recommend Larry's "math and music" page at: http://www.math.armstrong.edu/faculty/lesser/Mathemusician.html

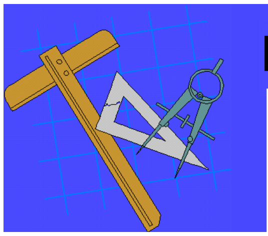
"AMERICAN π" by Lawrence Lesser
(reprinted with permission)

CHORUS: Find, find the value of pi, starts 3 point 1 4 1 5 9.
Good ol' boys gave it a try, but the decimal never dies,
The decimal never dies………

In the Hebrew Bible we do see
the circle ratio appears as three,
And the Rhind Papyrus does report four-thirds to the fourth,
& 22 sevenths Archimedes found
with polygons was a good upper bound.
The Chinese got it really keen:
three-five-five over one thirteen!
More joined the action
with arctan series and continued fractions.
In the seventeen-hundreds, my oh my,
the English coined the symbol π,
Then Lambert showed it was a lie
to look for rational π.
He started singing ………. (Repeat Chorus)

Late eighteen-hundreds, Lindemann shared
why a circle can't be squared
But there's no tellin' some people—
can't pop their bubble with Buffon's needle,
Like the country doctor who sought renown
from a new "truth" he thought he found.
The Indiana Senate floor
read his bill that made π four.
That bill got through the House
with a vote unanimous!
But in the end the statesmen sighed,
"It's not for us to decide,"
So the bill was left to die
Like the quest for rational π.
They started singing ………(Repeat Chorus)

That doctor's π in the sky dreams
may not look so extreme
If you take a look back: math'maticians long thought that
Deductive systems could be complete
and there was one true geometry.
Now in these computer times,
we test the best machines to find
π to a trillion places
that so far lack pattern's traces.
It's great when we can truly see
math as human history—
That adds curiosity……easy as π!
Let's all try singing……(Repeat Chorus)



Math Strategies

Group Folding and Groups Unfolded

A Gathering for Gardner IV, Atlanta, 2000

by Andy Liu†

We present an activity that is very suitable in a group setting. We call it *Scientific Origami*, as opposed to the usual *Artistic Origami* where one folds a piece of paper by following very complicated instructions, culminating in a beautiful bird, flower, or some other design.

We start with a square piece of paper that is blank on one side and coloured on the other. The usual origami paper is ideal for this purpose. In our illustrations, the coloured side will be green.

The instructions are very simple. We first fold the paper in half in one direction, and then in half again. Unfold the piece of paper and repeat the above steps in the other direction. We have creased the piece of paper in such a way that it is divided into 16 cells in a 4×4 configuration, as shown in Figure 1.

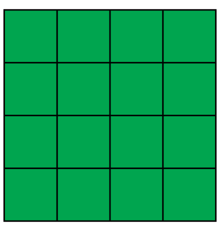


Figure 1

Our objective now is to fold this creased piece of paper to eliminate one row and one column. In the resulting 3×3 configuration, each of the nine cells is either completely blank or completely coloured.

It does not sound very complicated, does it? Figure 2 shows one way of doing it. After the two folds, we have a 3×3 configuration in which all nine cells are blank.

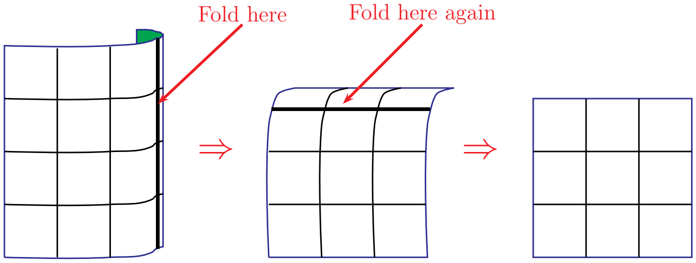


Figure 2

A design with nine green cells is equivalent to the one with no green cells. We can create it simply by turning over the piece of paper before we begin. From now on, we will regard such a pair of designs as a single one, and will use at most four green cells to represent it.

Looking at the back of the folded packet, we find a different design, with the nine squares divided 5:4, that is, with five coloured and four blank. Had we made the second fold in Figure 2 towards the front instead, we would have obtained two other designs, where the divisions are 7:2 and 6:3 respectively.

Exercise 1.

Fold a design where the squares are divided 8:1.

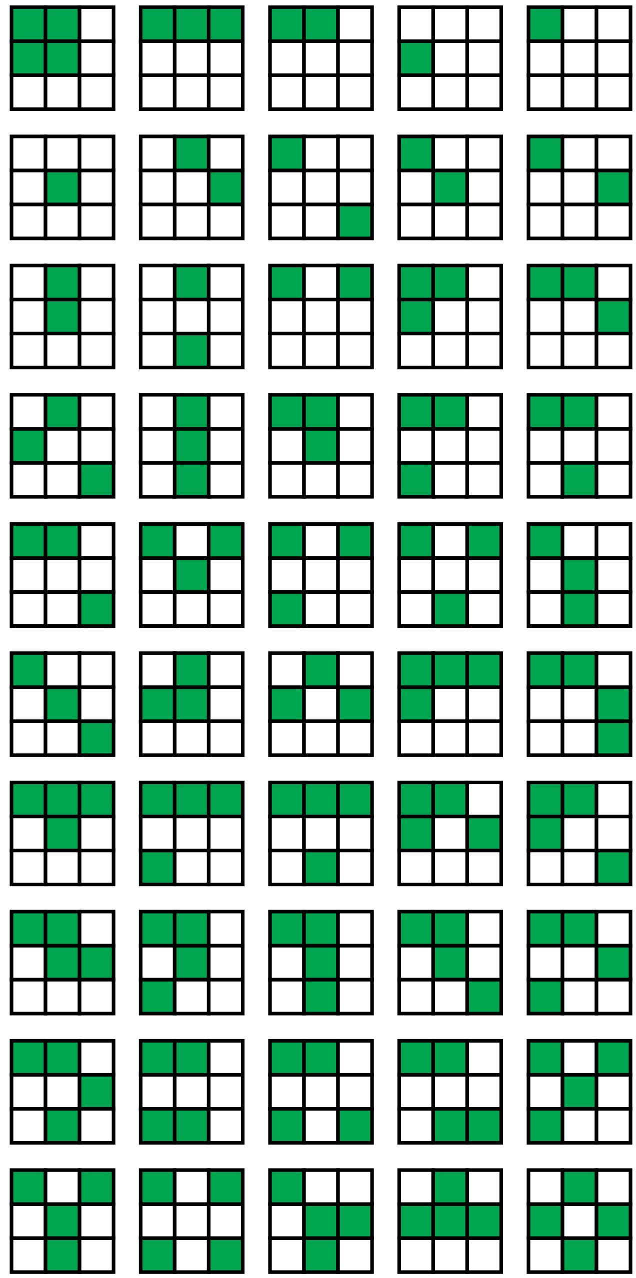


Figure 3

Exercise 2.

Fold as many as possible of the 50 designs in Figure 3. You should be warned that only about one-quarter of them can be achieved.

† **Andy Liu** is a professor in the Department of Mathematical and Statistical Sciences at the University of Alberta. His E-mail address is aliu@math.ualberta.ca.

The designs in Figure 3 plus the one in Figure 2 represent all there are. It may take a little while to check that they are all different; that is, they cannot be transformed into one another by rotation or reflection. The more difficult question is how we can tell that there are no more.

Let us first solve a related and simpler problem, in which the square is only divided into four cells in a 2×2 configuration. Here, we have $2^4 = 16$ distinct designs. Since it does not have a central cell, we will treat two designs as distinct even if they may be obtained from each other by changing blank cells to coloured ones and *vice versa*.

Two designs, with all four cells alike, stand alone and become patterns in their own right. Two others, with two diagonally opposite cells blank and the other two coloured, combine into a single pattern. Four other designs, consisting of two adjacent blank cells and two adjacent coloured ones, also form a pattern. Another four, consisting of one blank cell and three coloured cells, form a fifth pattern, while the last four designs form a sixth. Thus there are six distinct patterns.

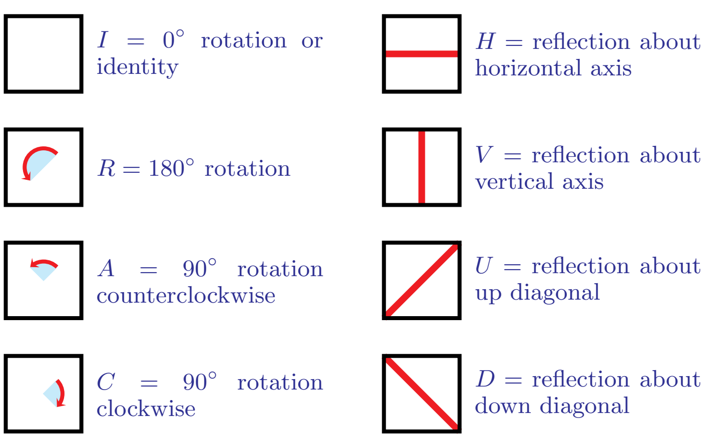
However, this is counting the hard way. While we can be reasonably satisfied that we have not missed out anything here, such confidence would be misplaced in the original problem of counting 3×3 patterns. We seek an alternative approach in which the task does not become significantly more difficult when the size of the problem increases.

Before getting into the more technical part of this article, we must acknowledge the inventors of this activity. Originally, Serhiy Grabarchuk of Ukraine asked for the folding of the 26th design in Figure 3 (three in a diagonal) from a 5×5 piece of paper. Later, three Japanese puzzlists by the names of Hiroshi, Kitajima, and Saseki, extended the puzzle and asked for the folding of all 51 designs from a piece of 5×5 paper.

Exercise 3.

Starting with a piece of 5×5 paper, fold all 51 designs in Figures 2 and 3.

Let us take a closer look at the transformations that bring a square back to itself. There are eight such symmetries, and they are listed in the chart below.



We now introduce a simple but very useful concept. A design is said to be *invariant* under a symmetry if the same design results after performing the transformation. Clearly, every design is invariant under I, but some designs are invariant under other transformations too. As an example, Figure 4 shows the action of the eight symmetries on four designs that combine to form a single pattern.

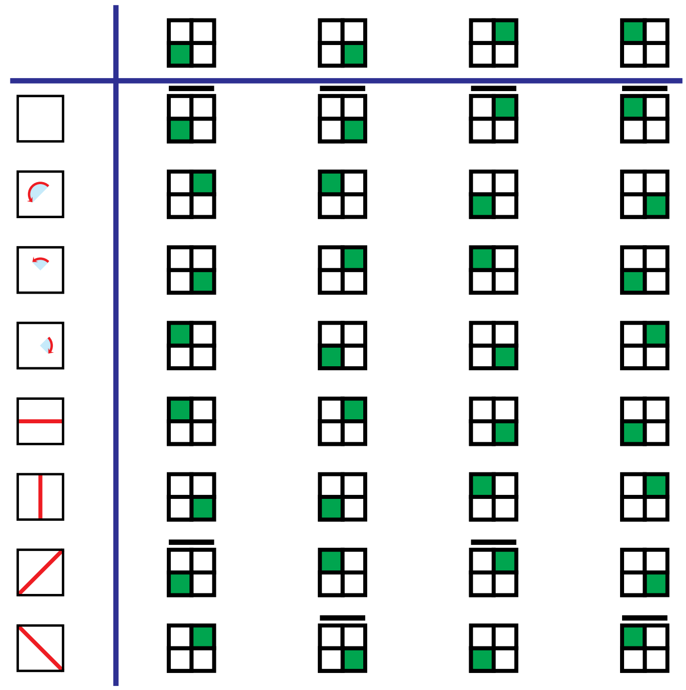


Figure 4

The invariant entries are marked with a bar on top of the square. There are eight such entries in this pattern. It is not difficult to verify that in each of the other five patterns, the total number of invariant designs under the symmetries is also eight. We claim that this is always the case, so that we can calculate the total number of patterns by dividing the total number of invariant entries by eight.

We now have to count invariant entries. It is just as difficult as counting patterns if we do it design by design. However, it turns out to be a much simpler task if we do it symmetry by symmetry.

Clearly, all $2^4 = 16$ designs are invariant under I. If a design is to be invariant under R, opposite cells must have the same colour. Since there are two pairs of opposite cells, the number of designs invariant under R is $2^2 = 4$. If a design is to be invariant under C, all four cells must have the same colour, so that the number of such designs is $2^1 = 2$. Similarly, the number of designs invariant under A is also $2^1 = 2$.

If a design is to be invariant under H, cells adjacent vertically must have the same colour. Since there are two pairs of such cells, the number of designs invariant under H is $2^2 = 4$. Similarly, so is the number of designs invariant under V. If a design is to be invariant under U, the two cells not on the up diagonal must have the same colour. Thus, we are free to choose the colours of three cells, so that the number of designs invariant under U is $2^3 = 8$. Similarly, so is the number of designs invariant under D.

It follows that the total number of invariant entries is 16+4+2+2+4+4+8+8=48. Dividing by 8, we obtain 6, confirming our earlier direct count. In the original problem, the number of invariant entries can be counted in a similar way to yield $2^9 + 2^5 + 2^3 + 2^3 + 2^6 + 2^6 + 2^6 + 2^6 = 816$, so that the number of distinct patterns is $816 \div 8 = 102$. Dividing by

2 to account for the fact that we do not distinguish colour-contrast, we indeed have 51 designs.

We still have to justify our claim that the number of invariant entries in each pattern is always eight. To do so, we have to introduce another concept, that of a **group**.

When we perform one symmetry of the square after another, the net result may be obtained by performing a single symmetry. For example, Figure 5 shows that if we first perform H and then A, the net result is U. We write $HA = U$.

Figure 5

Thus, we may define an operation of "followed by" among the symmetries of the square. The set $\mathcal{S}$ of symmetries of the square under this operation has the following properties:

1. **Closure:** Whenever X and Y are in $\mathcal{S}$, so is XY.
2. **Associativity:** For any X, Y and Z in $\mathcal{S}$, $X(YZ) = (XY)Z$.
3. **Identity:** There exists an I in $\mathcal{S}$ such that $XI = X = IX$ for any X in $\mathcal{S}$.
4. **Inverse:** For any X in $\mathcal{S}$, there exists a Y in $\mathcal{S}$ such that $XY = I = YX$.

Such a structure is called a *group*. In our example, the closure and identity properties are obvious. I indeed serves as the identity. C and A are inverse of each other while each of the other six is the inverse of itself. The associativity property is also clear since both $(XY)Z$ and $X(YZ)$ represent the net result of performing X, Y and Z in succession. The complete operation table is shown below.

First Symmetry	Second Symmetry							
	I	R	A	C	H	V	U	D
I	I	R	A	C	H	V	U	D
R	R	I	C	A	V	H	D	U
A	A	C	R	I	D	U	H	V
C	C	A	I	R	U	D	V	H
H	H	V	U	D	I	R	A	C
V	V	H	D	U	R	I	C	A
U	U	D	V	H	C	A	I	R
D	D	U	H	V	A	C	R	I

Note that the cancellation law holds, in that if $XY = XZ$, then $Y = Z$. Let W be the inverse of X. Then $W(XY) = W(XZ)$. By the associativity property, we have $(WX)Y = (WX)Z$. By the inverse property, we have $IY = IZ$. Finally, by the identity property, we have $Y = Z$ as claimed. This means that in each row of the table above, all eight entries are distinct. The same also holds for each column.

Two designs belong to the same pattern if and only if there is a symmetry that takes one to the other. From any one design, each of the eight symmetries takes it either back to itself or to one of the other designs in the same pattern. We illustrate this with the same pattern considered earlier, starting from the design at the bottom left corner of Figure 6.

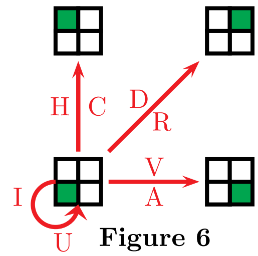


Figure 6

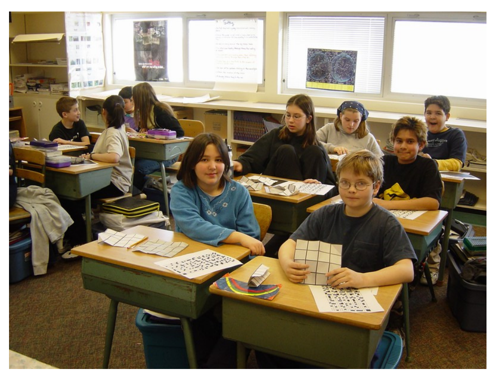
The starting design is invariant under two symmetries, namely, I and U. We claim that the number of symmetries under which another design is invariant is equal to the number of symmetries going to it from the starting design. Since each symmetry takes the starting design either back to itself or to another design, the total number of invariants within the pattern must be eight.

Consider, for example, the design at the bottom right corner of Figure 6. One of the symmetries that takes it back to the starting design is C. Then the symmetry CX takes this design to the same destination as the symmetry X takes the starting design. Moreover, the symmetries CI, CR, CA, CC, CH, CV, CU, and CD are distinct, being C, A, I, R, U, D, V, and H respectively. It follows that this design is invariant under $CA = I$ and $CV = D$, and under only these two symmetries.

This justifies our claim. The fact that the invariants are divided two apiece is immaterial, though it would have been most surprising were it not the case.

To conclude this article, we return to the easier problem of counting 2×2 patterns, not distinguishing colour contrast. All we have to do is to define eight additional symmetries, namely, I', R', A', C', H', V', D', and U', where X' means X followed by colour-reversal. It is easy to verify that this expanded set of symmetries form a 16-element group.

The numbers of invariants for the original eight symmetries remain unchanged. Those for the new symmetries are 0 for I', D' and U', 2 for A', and C', and 4 for R', H', and V'. Hence the total number of distinct patterns is given by $(48 + 16) \div 16 = 4$, as we have observed before.



Grade 6 students at Lynnwood Elementary School in Edmonton work on *Scientific Origami*.

Comment: This was the text of a talk to a group of Latvian youngsters on the occasion of an awards presentation for their National Mathematical Olympiad in Riga on May 30, 1999. The ceremony was presided over by the chief organizer, **Agnis Andjans**. He is a recent winner of the **Paul Erdős National Award** for the promotion of mathematics through competitions in his country. The award was bestowed by the **World Federation of National Mathematics Competitions**, founded by the late **Peter O'Halloran** of Australia.

Math Challenges

Problem 1. For how many positive integers x does there exist a positive integer y with $\frac{xy}{x+y} = 100$?

Problem 2. Find the number of non-negative integers n such that $2003 + n$ is a multiple of $n + 1$.

Problem 3. A cube of unit edge is rotated 30 degrees about one of its diagonals. What is the volume of the solid that is the intersection of the initial cube and the rotated one?

Send your solutions to π *in the Sky*: Math Challenges.

Solutions to the Problems Published in the September, 2002 Issue of π *in the Sky*:

Problem 1. Assume that $A \neq \emptyset$, where $\emptyset$ stands for the empty set. Then A contains at least one positive number. Indeed, if $x \in A$ and $x < 0$ then $x^2 - 5|x| + 9 \in A$ and

$$x^2 - 5|x| + 9 = |x|^2 - 5|x| + 9 = \left(|x| - \frac{5}{2}\right)^2 + \frac{11}{4} > 0.$$

Since $f(3) = 3$, then 3 could be an element of A. Also $f(x) = 3 \iff x \in \{\pm 2, \pm 3\}$. The only positive numbers in A could be 2 and 3. Indeed, assume by contradiction that there is a number $x \in A$, $x \neq 0$, $x \notin \{2,3\}$. Since $x \neq 3$, then $f(x) > x$. Now, since $f(x) \notin \{2,3\}$, we find that $f\big(f(x)\big) > f(x)$ and thus $f\big(f(x)\big) > f(x) > x$. Repeating this argument, we obtain an infinite sequence $x, f(x), f\big(f(x)\big), \ldots$ of distinct numbers that belong to A, thus A cannot be finite—a contradiction. We also mention that for any x, $f(x) \neq 2$. Hence if $x \in A$, $f(x)$ must be 3; therefore 3 must be an element of A. Let $x \in A$, $x \leq 0$. Since $f(x) \in A$ and $f(x) > 0$, we see that $f(x) = 3$; that is, $x \in \{-2, -3\}$. Therefore, we can take

$$\begin{array}{llll} A = \emptyset, & A = \{3\}, & A = \{-3,3\}, & A = \{2,3\}, \\ & A = \{-2,3\}, & A = \{-2,2,3\}, & A = \{-3,2,3\}, \\ & A = \{-3,-2,3\}, & A = \{-3,-2,2,3\}. & \end{array}$$

Problem 2. We can easily see that $(m,n) = (1,1)$, $(m,n) = (2,3)$ are solutions to the equation. Let us prove that other solutions do not exist. Suppose that there is another solution (m,n) not equal to those above. The equation can be written as follows:

$$3^m - 1 = 2^n \iff (3-1)(3^{m-1} + 3^{m-2} + \ldots 3 + 1) = 2^n$$
$$\iff 3^{m-1} + 3^{m-2} + \cdots + 3 + 1 = 2^{n-1}.$$

Since $n > 1$, the right-hand side is even. Hence, the left-hand side must be even as well, so that m must be even, say, $m = 2m'$. Since $m \neq 2$, we must have $m \geq 4$ and hence $n > 3$. Now we can write our equation as

$$9^{m'} - 1 = 2^n \quad \text{or} \quad (9-1)(9^{m'-1} + 9^{m'-2} + \cdots + 9 + 1) = 2^n,$$

or $9^{m'-1} + \cdots 9 + 1 = 2^{n-3}$. But $n > 3$, so the right-hand side, and also the left-hand side, of the above equality must be even. Thus $m' = 2m''$ and therefore $81^{m''} - 1 = 2^n$. However, this equality can not be true since the left side is divisible by 5, while the right side is not. Therefore, our assumption that we could have another solution has dropped.

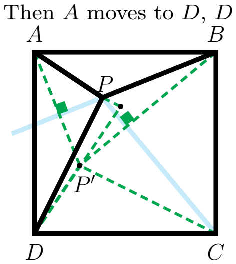
Problem 3. We may assume that $(m,n) = 1$. Use induction on m. For $m = 1$, the statement is clearly true. Let us assume that it is true for all fractions $\frac{k}{n}$ with $0 < \frac{k}{n} < 1$ and $k < m$. We prove that the statement is true for $k = m$, where $0 < \frac{m}{n} < 1$. By the division algorithm, we have $n = mq + r$, $0 < r < m$; that is, $m(q+1) = n + (m-r)$ or $m(q+1) = n + p$, for $0 < p < m$. This equality can be written as $\frac{m}{n} = \frac{1}{q+1}\left(1 + \frac{p}{n}\right)$. Using the induction hypothesis, $\frac{p}{n} = \frac{1}{b_1} + \cdots + \frac{1}{b_{s-1}}$, and b_{r-1} divides b_r, $r = 2, s-1$, hence

$$\frac{m}{n} = \frac{1}{q+1} + \frac{1}{(q+1)b_1} + \cdots + \frac{1}{(q+1)b_{s-1}},$$

and the statement holds for $m = k$ if $a_1 = q+1$, $a_2 = (q+1)b_1, \cdots$, $a_s = (q+1)b_{s-1}$.

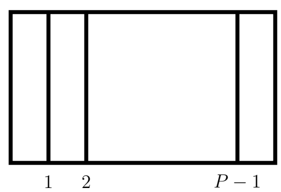
Note: The above solution provides an algorithm for writing a fraction as a sum of fractions with numerator 1. For example, if $\frac{7}{11}$, then $2 \cdot 7 = 11 + 3 \Longrightarrow \frac{7}{11} = \frac{1}{2}\left(1 + \frac{3}{11}\right)$, $3 \cdot 4 = 11 + 1 \Longrightarrow \frac{3}{11} = \frac{1}{4}\left(1 + \frac{1}{11}\right)$, hence $\frac{7}{11} = \frac{1}{2}\left(1 + \frac{1}{4}\left(1 + \frac{1}{11}\right)\right) = \frac{1}{2} + \frac{1}{2 \cdot 4} + \frac{1}{2 \cdot 4 \cdot 11}$.

Problem 4. Rotate the square about its center O, counterclockwise $90°$. Then A moves to D, D to C, C to B, B to A, and P to some point P'.

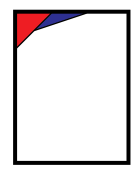


The lines AP, DP, CP, and BP move in four lines through D, C, B, and A respectively. These lines are perpendicular on AP, DP, CP, and BP respectively, and all intersect at P'. The perpendiculars from the problem are three of these lines; hence, they intercept at P'.

Problem 5. Making a new cut in a polygonal piece of cardboard, we increase the number of pieces by one. If we make N cuts, we get $N+1$ pieces. If we have a set of polygonal pieces of cardboard and cut one piece, the number of vertices of the new set will increase with at most four vertices. Since initially we have four vertices, after N cuts we cannot have more than $4N + 4$ vertices in all pieces. Let us assume that after N cuts we have got P polygons with S sides. Since we have a total of $N+1$ pieces, $N + 1 - P$ polygons do not have S sides. On the other hand, each piece (of these $N+1$ polygons) has at least three vertices, hence the number of vertices for all pieces is at least $P \cdot S + (N - P + 1)3$. Therefore, $PS + (N - P + 1)3 \leq 4N + 4$. That is, $PS - 3P - 1 \leq N$. This inequality says that, in order to obtain P polygons with S sides, we have to make at least $PS - 3P - 1$ cuts. Let us prove that this number is enough to achieve our goal.



With $P-1$ cuts, a rectangle is tranformed into P rectangles.



Each rectangle can be transformed into a polygon with S sides by making $S-4$ cuts.

Thus, by making $(S-4) \cdot P + P - 1 = PS - 3P - 1$ cuts, we can obtain P polygons with S sides. In our problem, we have $P = 2002$ and $S = 2003$. Hence the minimum number of cuts is $2002 \cdot 2003 - 6006 - 1 = 4003999$.

Problem 6. For $k > 0$ and A, B real numbers we have the inequality $AB \leq \frac{1}{4k}(A + kB)^2$. Taking $A = \sum_{i=1}^n p_i f(x_i)$, $B = \sum_{i=1}^n p_i g(x_i)$, we get

$$\left(\sum_{i=1}^n p_i f(x_i)\right)\left(\sum_{i=1}^n p_i g(x_i)\right) \leq \frac{1}{4k}\left(\sum_{i=1}^n p_i f(x_i) + k\sum_{i=1}^n p_i g(x_i)\right)^2$$
$$= \frac{1}{4k}\left(\sum_{i=1}^n p_i\left[f(x_i) + kg(x_i)\right]\right)^2. \quad (1)$$

The function $h(x) = f(x) + kg(x)$ is convex on $[a,b]$, hence $h(x) \leq \max\{h(a), h(b)\}$, for every $x \in [a,b]$ (see π *in the Sky*, September 2002, Math. Strategies). Hence $f(x_i) + kg(x_i) \leq \max\{f(a) + kg(a), f(b) + kg(b)\} = M$ and the required inequality follows from (1).

Note: If we take $f(x) = x$ and $g(x) = \frac{1}{x}$, $0 \leq a \leq x_i \leq b$, $p_i \geq 0$, for $i = 1, \ldots, n$ we get

$$\left(\sum_{i=1}^n p_i x_i\right)\left(\sum_{i=1}^n \frac{p_i}{x_i}\right) \leq \frac{(a+b)^2}{4ab}\left(\sum_{i=1}^n p_i\right)^2,$$

which is Kantorovich's Inequality (π *in the Sky*, September 2002, Math. Strategies).