Problem Set 1

Opener: "Fake It, Make It"

1. With a partner or two, *fake* the results of flipping a coin 360 times. The goal is to make the most believable fake possible, one that could fool others into thinking the flips came from a real coin. Write heads as 1, tails as 0. Don't use anything other than pencil and paper.

Each problem set opens with a game to play or analyze. After that, the problems are divided into Important Stuff, Neat Stuff, and Tough Stuff.

2. Now, with the same group, *make* the results of flipping a coin 360 times. Write heads as 1, tails as 0. Do not use technology or other aids. Just flip a coin 360 times and write down what it says. Identify the two lists in some way that you would recognize, but a neighbor would not.

3.
- **a.** Create a test you could use to decide whether a list someone gave you was real or fake.
- **b.** Now exchange lists with a neighboring group and run your test. Was the test successful?

Important Stuff problems are key in the overall course. They're important! Do these after the opening game.

Important Stuff

4.
- **a.** Find the probability that when you pick two integers between 1 and 5 (inclusive), they do not share a common factor greater than 1.
- **b.** Repeat for picking between 1 and 6, 1 and 7, 1 and 8, 1 and 9.

Keep track of the assumptions you make about picking the integers. There is more than one way to pick. It is interesting to compare the different options here, and the differences in corresponding results.

5.
- **a.** If you flip a fair coin 360 times, how many heads would you expect?
- **b.** Take a guess: what is the probability of getting *exactly* this many heads?

Take quick guesses. 10 seconds or less, please! You will revisit these questions later.

6.
- **a.** If you roll a fair die 360 times, how many times would you expect to roll a one?
- **b.** Another guess: is it more likely to get exactly this many ones, or to get exactly the number of heads from Problem 5?

Neat Stuff

Neat Stuff problems are interesting but may not be on the "main track". Feel free to pick and choose, but keep in mind these problems may become important later!

7. What's the probability that an integer picked from 1 to n is a perfect square if
- **a.** $n = 20$?
- **b.** $n = 200$?
- **c.** $n = 2000$?
- **d.** $n = 20000$?

e. What is happening "in the long run" (as n grows larger without bound)?

8. Patsy offers you these two games:

Game 1: You roll a die four times. If you roll a six any of the four times, you win.

Game 2: You roll a pair of dice 24 times. If you roll boxcars (double sixes) any of the 24 times, you win.

Aside from the fact that Game 2 takes longer to play, which of these games would you rather play to win? Or do both games have the same chance of winning?

Problem 8 lies at the foundation of probability theory, and was originally solved by Pascal.

9. The new toy craze is Mega Men, where kids buy a Mega Man in a box without looking to see which one it is, then open it up when they get home. There are ten toys in all, each equally likely when you buy a box. Kids want to collect all ten, but only a new-in-box toy will do.

On average, how many boxes will Therese have to buy for her son before he can finally collect them all?

The answer isn't 10! There will probably be plenty of repeats. . .

Tough Stuff

Tough Stuff problems are hard; sometimes very hard; sometimes impossible! They are fun to play around with, though.

10. In Yahtzee, you get three rolls and you're looking to get all 5 dice to be the same number. You can "save" dice from one roll to the next. There are other goals, but this problem explores the 5-dice Yahtzee.

a. Find the probability that if you try for it, you will get a Yahtzee of all 6s by the end of your third roll.

b. (*harder*) Find the probability that if you try for it, you will get some Yahtzee by the end of your third roll. Assume that you always play toward the nearest available Yahtzee: if your first roll is 2-3-3-5-6, keep the 3s and roll the rest.

Problem Set 2

Opener: "Bingo Hi-Lo"

On an unpopular game show, a "base number" between 1 and 75 is randomly picked. A contestant is then asked whether the next number (picked from all *remaining* numbers) will be higher or lower. If they are correct, the contestant wins $100 multiplied by the number that comes out. Suppose you are the contestant and you're trying to win as much money as possible.

Suppose the base number is 25 and you said "Higher". If the next number is 35, you win $3500, but if it is 17, you win nothing. And yes, only integers are picked!

1. **a.** If the base number is 30, would you choose higher or lower? Why?
b. If the base number is 38, would you choose higher or lower? Why?
c. If the base number is 45, would you choose higher or lower? Why?
d. State an overall strategy as clearly as possible, and be prepared to defend it to others!

Important Stuff

2. Describe some ways to calculate the sum

$$1 + 2 + 3 + 4 + \cdots + 74 + 75$$

3. Describe, in complete detail, a test you could perform on a set of 360 coin flips that would help you decide whether it is real or fake. Try to revise or improve your test from Problem Set 1.

4. *Use your test* on your choice of four of these data sets to decide whether they are real or fake.

In each of these, consecutive flips are adjacent horizontally. Set (a) begins with a tail, then 5 heads in a row.

a.

0111110001	1100111111	1100000000	0000001010
1010101010	1010101010	1010010001	0010101010
1011111100	0000111111	0000001010	1010101000
0010101010	1010000001	0100001010	0001010001
0101001010	1001010100	1010101010	1001010101
0001001000	1010001010	1001000001	0000010000
0100000000	1000001000	0010001010	1010101010
1010101010	1010101000	0001110000	0001010100
0101010100	0100100010	0100010010	0000001101

b.

0101000100	0101010110	1011010101	0101010100
1000101010	0010011101	1011011010	1010101010
1001010101	0101000001	0101010101	0101000101
0101010110	1010101010	1001010101	0010100101
0101010101	0101010111	1011010010	0101010101
0100100111	1010101010	1010100101	0101011101
0101010010	0010101010	1010100010	1110101011
1010101101	0101000101	0101111010	1010101000
0101010100	0001110101	0111011101	0110101110

c.

1101101001	1001111011	1101001000	1000011110
1101010100	1001111111	1100011101	0010000011
0010110011	1000010110	1011000010	1011110010
1010100000	0011010101	0010010001	0110110110
1111111110	0011001001	0101100100	1011000100
1010000000	1010010100	0101001011	0010010110
0110000001	1110100011	0111010100	1111100100
0001100001	0101010110	1110000010	1100011000
0000101110	1010011001	0001100010	0011111101

d.

1000100000	0000000001	0010010100	1101010101
1010110111	0000101011	1111100000	1100001010
0010001011	0011101111	0101111010	0010011000
1100000010	0000100110	0100100100	1110110011
1000101011	0100111101	0110101110	0110011000
1101011011	0001000110	0011110000	0000110011
1101010001	0110100010	1111101101	1110000101
0011011111	0110110011	1010110011	1011100101
0010000100	0100010100	0111011000	0110101000

e.

0011000110	1010011101	1100101001	1100011100
0110110101	0111001010	1011010101	0001101010
0110101000	1010101011	0110100011	1101111011
0101010100	1101110110	1110001010	1011101010
1011101111	0101001010	1000011001	1011110110
1011010111	0110100100	0100100101	0100011010
1010110101	0100101110	1010111010	0101011011
1101010100	1010101010	0010101010	1010100101
0100100101	0101101010	1010010101	0101010111

f.

0010011001	1010101010	0100101110	0100110010
0101001101	0101010100	1001010100	1110011001
1010011001	0010010100	1100110101	1001100011
0010111010	0100100101	0110001100	1001101010
1010101010	1010101100	1001100110	0011001110
0110011001	1010101101	0001110011	0010011010
1100101001	0101010100	1011001010	1010101001
0101010101	0100110101	0010010101	0101001100
1011010101	0101010100	1100100110	0110011010

g.

1100001111	0010010011	1000011000	1100001011
0101001110	1101000011	1111111000	1001000111
0111010011	1110101010	0010101010	1010010101
1000000001	0010000010	1001100110	1011101011
1111100001	0011111111	1011010010	1101101101
1000110110	1100010101	0011101100	1111101010
0111110111	1111110001	0110101001	1101011101
0100001000	0100110110	1001111111	0100000011
0000011100	0111110000	0101111110	1011001001

5. **a.** Find the probability that if you pick two integers between 1 and 6 (inclusive), the two numbers picked do not share a common factor greater than 1. There is more than one possible correct answer.

b. Repeat for picking between 1 and 7, 1 and 8, 1 and 9, 1 and 10. Are any patterns emerging?

6. Work on Problems 5 and 6 from Problem Set 1 if you haven't already.

7. Jocelyn has a piece of paper. She tears it into three equal pieces and hands one piece to Dave, another to Keith, and keeps the third piece for herself.

She continues to do this; tearing the paper she has left into three equal pieces, handing one piece to Dave, one to Keith, and keeping the third.

a. After two tears, how much paper does Jocelyn have left? How much do Dave and Keith each have?

b. After three tears?

c. After four tears?

d. After 10 tears?

e. Forever?

f. Write two different expressions for the amount of paper Dave has after this is over.

8. Mary has a piece of paper. She tears it into four equal pieces and hands one piece to Manuel, one piece to Kim, one piece to Sandra, and keeps the fourth piece for herself.

She continues to do this.

a. After two tears, how much paper does Mary have left? How much do Manuel, Kim, and Sandra each have?
b. After three tears?
c. After four tears?
d. After 10 tears?
e. Forever?
f. Write two different expressions for the amount of paper Sandra has after this is over.

Neat Stuff

9. What's the probability that an integer picked from 1 to n is *not* a perfect square if

a. $n = 25$?
b. $n = 250$?
c. $n = 2500$?
d. $n = 25000$?
e. What is happening "in the long run" (as n grows larger without bound)?

10. Roll a six-sided die and record the result. Keep rolling until you've rolled each number from 1 to 6 at least once. What is the expected number of rolls you will need before achieving this?

Unless otherwise noted, assume that a die or dice has the standard numbers 1 through 6 on its sides.

11. In a carnival game, you win if you roll dice that add up to 15. But, before rolling, you must choose the exact number of dice to roll. What number of dice gives you the best chance of rolling a sum of 15?

12. In an unusual coin-flipping game, you get 1 point every time you flip heads. But, flip tails and you're in "danger" and must flip heads next. If you flip tails twice in a row, you "bust" and lose all your points (but continue playing). The game lasts 10 flips.

a. Find the probability that you survive all 10 flips without busting even once.

b. What is the average score you could expect after 10 flips?

c. What would happen in a longer game? Will the average score increase or decrease? Is there a limit?

Tough Stuff

Feel free to revisit problems from previous problem sets at any time.

13. In Problem 9 of Problem Set 1, you looked at the expected number of Mega Man toys you'd need to get one of each of the ten models. But what about more than one of each?

Ilene's two kids will fight like crazy unless they *each* have one of every Mega Man. On average, how many boxes will you have to buy in order to have at least *two* of each model?

14. Build a data set with at least 5 elements such that if m is the mean and n is the median, then $|m - n|$ is larger than the standard deviation of the set.

Use the Internet or some other reference to learn how to calculate standard deviation, or use the `stDevPop` function on the TI-Nspire.

CHAPTER 2 Facilitator Notes

The facilitator notes are designed to be used as needed. Each problem set has two components:

1. **Goals of the Problem Set** Here we lay out what the principal ideas of each problem set are.
2. **Notes on Selected Problems** We identify a few problems that are worth going over in a whole group discussion.

We will put our emphasis on the main goals of each lesson, drawn from the problems in the "Important Stuff."

An overall course note: this course uses some applications of CAS (computer algebra system) technology. Workarounds are provided in the detailed notes on each problem set, but the best tool is the TI-Nspire CAS calculator, for which detailed instructions are provided within the problem sets. Another good option is the TI-Nspire CAS Teacher's Edition, an emulator for PC or Mac. There are many other options, including TI-89, other calculators, Mathematica, Maple, MATLAB, and a web option, Wolfram Alpha. Some of the material later in this course is difficult (actually, more tedious than difficult) to work through without a CAS, so it is strongly recommended but not required.

Problem Set 1

Goals of the Problem Set

This course is about some connections between algebra and probability. By the end of Problem Set 14, participants will construct algebraic models for repeated experiments such as coin flips and dice rolls, and ways of determining the mean, variance, and standard deviation for any number of such experiments.

The course begins simply, since there is little or no assumed knowledge of probability and statistics. This problem set's opener is one that will be followed up several times, and shows the difficulty of trying to "fake" randomness.

Problem Set 1 is also intended as an introduction to the style of the course. Consider having participants read the Introduction to learn about the course expectations. Remember that participants are not expected to move through the entire problem set, and many will not move past "Important Stuff". That's fine; things in "Neat Stuff" are generally extensions or previews.

Notes on the Problems

The opener will take a long time. Make clear to participants that they are building *ordered* lists of coin flips as 1s and 0s, as this may be unclear. Watch out for participants flipping multiple coins at once in Problem 2, since they may be writing down these results in an order that removes the randomness – at PCMI 2007 one group flipped ten coins at once, then wrote down all the 1s for heads and all the 0s for tails, which gave long runs of 1s and 0s that were invalid results. If anything like this happens, try to catch it quickly, since the experiment needs to be restarted and takes a while.

If time is an issue, consider shortening the number of flips to 120 or 240. We chose 360 since it is highly factorable; it divides evenly by 1 through 6, 8, 10, 12, and more. This allows the 360 flips to be "grouped" more easily. Participants may prefer to use graph paper since it will allow the flip lists to be more easily organized into groups of 10 or 20 flips.

For Problem 3, one quick test you can use yourself (but don't tell) is to look for the longest "run" of consecutive heads or tails. Truly random data is very, very likely to have a run of at least 8 consecutive heads or tails, while constructed data usually does not have such a run. One interesting test participants constructed was to count the total number of "changes" (from head to tail, or tail to head). Truly random data would have about 180 "changes" since they should occur half the time. Faked data typically has significantly more "changes".

There are many open questions to introduce here. While 180 heads, or 180 "changes", is expected, it won't likely be exactly 180 of anything. So how many is too many or too few? 190? 200? 210? more? How long should the longest "run" be? Don't answer any of these questions here, but consider returning to them in later problem sets. Problems 5 and 6 are meant to tease out some of these concepts.

Problem 4 is the beginning of a long track about "relatively prime" integers. Expect some disagreement here, since there are several ways to pick two integers between 1 and 5! Is $(5, 1)$ the same as $(1, 5)$? Is $(2, 2)$ allowed? The sidenote to this problem notes the fact that there will be disagreement. Later problem sets (Problem Set 4, Problem 6) will take a more directed approach to this problem, specifically that $x \geqslant y$ is required. The end goal of the track is to measure the proportion of pairs that share no common factors, and it turns out that regardless of which approach is taken, the limit of this proportion will be the same. (No need to bring that up in this problem set, though.) What is important here is that in many probability and statistics problems, the statement of the problem needs to be clear in order to avoid this sort of ambiguity.

Problem Set 2

Goals of the Problem Set

This problem set continues the thread from Problem Set 1 regarding the "coin test", to help develop participants' understanding of some characteristics of randomness.

Some of the questions in this problem set also target a specific long-term goal, which is to answer this question: What is the probability that two numbers picked at random are relatively prime? This problem set seeks answers to this question, up to numbers chosen between 1 and 10, while the long-term goal is for this question to be answered with no limit to the numbers involved.

As with Problem Set 1, you may expect some wide gaps in knowledge and familiarity with probability concepts, and similar gaps in how far participants proceed through the problem set. Discussions should center on problems in "Important Stuff" unless all participants have advanced.

Notes on the Problems

While parts (a) and (b) of the opener are straightforward, part (c) is likely to generate debate. Most participants will initially choose "lower" simply because it is more likely to win, but the amount of money won is relevant. It turns out that more money (on average) is won by choosing "higher", even though the chance of winning at all is below 50%. Consider giving a careful analysis for 45, then asking participants to determine where a useful "cutoff" would be. This cutoff turns out to be much higher than expected.

Problem 2 is a followup to the opener, but turns out to be useful in a second way: two of the strategies participants may use in Problem 5 (and in Problem 4 of Problem Set 1) count the number of elements in this way. The "staircase" method is in Problem 6 of Problem Set 3, so unless that comes up, don't show it.

In Problem 5, try to collect patterns or observations from groups using different counting strategies. A nice goal here is to show that the results of the three different counting strategies (any values, values where $x \geqslant y$ is required, values where $x > y$ is required) can be compared. The numerators should be closely related – particularly, the numerator for "any values" is exactly double the equal numerators of the others. And the denominators can be determined quickly (as n^2 or by using a staircase). A second key observation is that the value of the fractions determined by this process are all approximately equal. In the end, one of these methods will be agreed upon and used, but it is important to recognize that this choice doesn't change the nature of the process and won't have a significant effect on the outcome of the calculations.

Problems 7 and 8 introduce an interesting way of justifying the rules for the sum of a geometric series. This formula will be useful in dealing with situations that don't have a clear conclusion, like the average number of heads before a tail appears when flipping a coin. Geometric series are rarely central in this course, so you may skip these problems if necessary.

CHAPTER

4 Mathematical Overview

This overview contains a sample of the mathematical themes that the development team hammered out in the process of designing the course. It contains some of the mathematical background used when creating the problem sets as well as some mathematical extensions that never made it into the "soup" of problems that went out to PCMI. It is written by people who were in on the design but were not involved in the day-to-day classes at PCMI.

In addition to the teachers participating in the course, PCMI hosts research programs in mathematics and education, programs for graduate students and undergraduate faculty, and institutes for staff development professionals. See `pcmi.ias.edu` for more details.

The original plan for the course was to weave four themes throughout the three weeks:

1. the use of expected value and variance in the analysis of games and related situations,
2. the use of formal algebra with polynomials as combinatorial tools,
3. an introduction to probabilistic number theory, and
4. an investigation into the distribution of primes in the integers.

This turned out to be too much material for three weeks, so the last item didn't make it into the summer course, although this theme was the subject of a great deal of discussion in the writing team.

Expected Value and Variance

These functions are developed gradually, in very concrete settings, over the course of the problem sets. Their properties are illustrated in contexts (often games), and, while the formal proofs are not made explicit, they are embedded in the problems—see the opener for Problem Set 5 or problems 10 and 11 in Set 9, for example. We'll list some of the properties here. Most of the proofs can be found in Chapter 7 of [1].

Notation and Definition

We'll think of a *sample space* as a set of atomic outcomes. Although the notation is not used in the problem sets, we'll let $P(A)$ be the probability measure of a subset A of S. For example, if S is a finite set and the outcomes in S are equally likely,

$$P(A) = \frac{|A|}{|S|}$$

where $|\cdot|$ is the number of elements in a set.

If S is a sample space, a random variable is a function $X : S \to \mathbb{R}$. In this course, S is always a finite set, so X is a function that takes on only a finite set of values. We'll use the notation $P(X = k)$ to mean the probability that X takes on the value k. For example, if X returns the number of heads when five fair coins are tossed, the values of X come from the set $\{0, 1, 2, 3, 4, 5\}$, and, for k in this set,

$$P(X = k) = \frac{1}{32}\binom{5}{k}$$

For example, if S is a set of throws of two dice, a random variable X might assign a given throw to the sum of the values on the faces, or it might return the product of these values, or the difference, or the sum of the squares . . . .

In most of the examples in the course, the values produced by X are non-negative integers. But they need not be.

The *expected value* of X is the average of the values of X, weighted by their probabilities:

$$E(X) = \sum_k kP(X = k)$$

In this course, this is a finite sum. An equivalent formula is $E(X) = \sum_{s \in S} P(s)X(s)$.

The definition is a mouthful, but the Plinko game is a good example: each k is the dollar value, and each $P(X = k)$ is the probability of hitting that value.

As another example, for the X in the five-coin toss above, we have:

Number of Heads	Probability	Product
0	1/32	0/32
1	5/32	5/32
2	10/32	20/32
3	10/32	30/32
4	5/32	20/32
5	1/32	5/32
	sum =	**80/32**

The expected value is $\frac{80}{32}$, which simplifies to $\frac{5}{2}$.

Does $\frac{5}{2}$ make sense as the average number of heads when flipping five coins?

Two more definitions:

- In the course, *variance* is viewed as a property of data sets— it is the "mean squared deviation:" If the data set is $\{x_1, \ldots, x_n\}$, the variance is

$$\sigma^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

where $\overline{x}$ is the mean of the x_i. More generally, if X is a random variable, its variance is the expected value of the square of the deviation from its expected value (another mouthlful):

$$\mathrm{Var}(X) := E\left((X - E(X))^2\right)$$

- The notation σ^2 is loaded: The standard deviation is the square root of the variance. For data sets, it says

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

and for a random variable, we have

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

"The mean of the squares of the deviation from the mean."

$X - E(X)$ is a random variable: the function defined by $s \mapsto X(s) - E(X)$. It differs from X by a constant function.

The main properties of expected value, variance, and standard deviation that are illustrated or developed in the course are:

Theorem 4.1

Suppose X and Y are random variables defined on the same sample space. Then

- $E(X+Y) = E(X) + E(Y)$
- *If* X *and* Y *are independent random variables, then*

$$Var(X+Y) = Var(X) + Var(Y)$$

The proofs are in [1].

Two random variables X and Y are *independent* if $P\,(X = k \text{ and } Y = j) = P(X = k)\cdot P(Y = j)$
If X and Y are independent random variables, then $E(XY) = E(X)\,E(Y)$.

If X and Y are not independent, we have an interesting calculation:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E\left((X+Y-E(X+Y))^2\right)\\
&= E\left((X-E(X)+Y-E(Y))^2\right)\\
&= E\left(\left((X-E(X))^2 + 2(X-E(X))(Y-E(Y)) + (Y-E(Y))^2\right)^2\right)\\
&= E\left(\left((X-E(X))^2\right) + 2E\left((X-E(X))(Y-E(Y))\right) + E\left((Y-E(Y))^2\right)\right)\\
&= \mathrm{Var}(X) + 2E\left((X-E(X))(Y-E(Y))\right) + \mathrm{Var}(Y)
\end{aligned}$$

So, the "obstruction" to the linearity of variance is that $E((X - E(X))(Y - E(Y)))$ might not be 0. This function of X and Y is called the *covariance*:

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

Equipped with this definition, we have a more general fact about the variance of a sum:

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$$

So, if two random variables are independent, their covariance is 0. *Note that the converse is false*: Two random variables can be dependent and still have 0 covariance. An example is if X takes on values $-1, 0, 1$ with equal probability and $Y = X^2$. Still, covariance can be used to measure dependence (if $\text{Cov}(X, Y) \neq 0$, X and Y are dependent).

There's a strong analogy between these constructs and results and ideas from linear algebra. Although not developed in the course, this analogy might be useful for facilitators.

- Because they are functions $S \to \mathbb{R}$, the set of random variables on a sample space form, together with addition and scaling, a *vector space* over $\mathbb{R}$. And linear dependence in this space implies dependence of the random variables in the probabilistic sense.
- Expected value is a linear map on this vector space.
- The way in which standard deviation is calculated looks like the distance formula, and, indeed, σ has the same properties as a length.
- Covariance is inner product on this space, and there's even is a Cauchy-Schwarz inequality:

  $$-\sigma(X)\,\sigma(Y) \leqslant \text{Cov}(X, Y) \leqslant \sigma(X)\,\sigma(Y)$$

  which implies that

  $$-1 \leqslant \frac{\text{Cov}(X, Y)}{\sigma(X)\,\sigma(Y)} \leqslant 1$$

The value of $\frac{\text{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)}$ is the *correlation coefficient* for X and Y, let's call it $\text{Cor}(X, Y)$, and Cauchy-Schwarz tells us that it's always between -1 and 1 (and that it's 0 if X and Y are independent). This looks just like the formula for the angle between two vectors.

Hence we have a structural similarity between our space Ω of random variables and $\mathbb{R}^n$

	$\mathbb{R}^n$	Ω
Vectors	n-tuples	random variables
Inner product	dot product	covariance
Length	$\|X\| = \sqrt{X \cdot X}$	$\sigma(X) = \sqrt{\text{Var}(X)}$
Cauchy-Schwarz	$-\|X\|\,\|Y\| \leqslant X \cdot Y \leqslant \|X\|\,\|Y\|$	$-\sigma(X)\,\sigma(Y) \leqslant \text{Cov}(X,Y) \leqslant \sigma(X)\,\sigma(Y)$
Angle between vectors	$\cos(X,Y) = \frac{X \cdot Y}{\|X\|\|Y\|}$	$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)}$

Finally, the problem sets introduce the "machine formula" for variance. This amounts to establishing an algebraic identity:

$$\frac{\sum(x-\overline{x})^2}{n} = \overline{x^2} - \overline{x}^2$$

Here's a simple proof:

$$\begin{aligned}
\frac{\sum(x-\overline{x})^2}{n} &= \frac{\sum(x^2 - 2x\overline{x} + \overline{x}^2)}{n} \\
&= \sum \frac{x^2}{n} - 2\sum \frac{x\overline{x}}{n} + \sum \frac{\overline{x}^2}{n} \\
&= \sum \frac{x^2}{n} - 2\overline{x}\sum \frac{x}{n} + \overline{x}^2 \sum \frac{1}{n} \\
&= \sum \frac{x^2}{n} - 2\overline{x}^2 + \overline{x}^2 \\
&= \sum \frac{x^2}{n} - \overline{x}^2 \\
&= \overline{x^2} - \overline{x}^2
\end{aligned}$$

This identity is reminiscent of the famous Lagrange identity:

$$n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{1 \leqslant i < j \leqslant n} (x_j - x_i)^2$$

And taken together, we have an interesting algebraic fact that you might ask students to derive:

This is a very intriguing identity—it replaces a sum of pairwise differences by a sum of single differences.

$$\frac{1}{n}\sum_{1 \leqslant i < j \leqslant n} (x_j - x_i)^2 = \sum_i (x_i - \overline{x})^2$$

Formal Polynomials

One way to use polynomials as modeling tools is to look at polynomials as objects that define functions. In this way, gravity can be modeled with a quadratic polynomial, and volume can often be modeled with a cubic one. When polynomials are viewed as functions, the "x" is thought of as a *variable*, a generic element of some replacement set.

There is another use of polynomials in which the "x" is an *indeterminate*. In this view of a polynomial, the letters are just placeholders (rather than "valueholders")—what's really important are the operations *between* the letters. The difference can be illustrated with two common activities in school algebra: simplifying and graphing. When students simplify polynomials, they are thinking of them as formal objects; the fact that $x^2 - 1 = (x-1)(x+1)$ comes from the fact that, if the right side is expanded by "the rules of algebra" you end up with the left side. Graphing, on the other hand, requires that you think of substituting values for x, or that you imagine x sweeping across some domain, producing points on a graph.

If two polynomials are equivalent formally, they define the same function. And there's an important partial converse to this: if two polynomial functions agree at "enough" inputs, one can be obtained from the other by formal calculation.

Of course, both of these points of view are important and not completely disjoint, and we often want students to be able to move between them, sometimes in the same problem. For example, in elementary algebra, it is often helpful to forget the fact that the letters stand for numbers, although we always know that they are, in fact, just placeholders for numbers.

There are times, however, when the *form* of a calculation is more important, and we never think of replacing the letters with values. For example, at several points in the course, participants are asked to investigate the distribution of possible sums when 3 (or more) dice are thrown. And they find the coefficient of x^9 when

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^3$$

is expanded. The object here is *not* to perform the expansion by hand or machine, but to calculate without calculating, reasoning where an x^9 term can occur when one multiplies out the expression. The formal calculation here models the counting problem for three dice, providing another way algebra can be used as a modeling tool. Such reasoning requires an example of what Common Core calls "decontextualization"—a formal approach to polynomial calculations that is important enough to deserve increased attention in the later years of high school.

The lines between form and value are not clean. By replacing x by -1 in the un-expanded form, students can show that the number of outcomes with an odd sum is the same as the number of even outcomes when any number of dice are thrown.

Example: Bowen's game

Bowen Kerins (one of the instructors at PCMI and an author of this course) poses the following game: Suppose you have a tetrahedral die with faces numbered $3, 4, 6, 9$. You can throw it as often as you want until you come up with a sum of 20. In how many ways can this happen?

The coefficient of x^{20} in $\left(x^3 + x^4 + x^6 + x^9\right)^n$ is the number of ways you can roll a sum of 20 with n throws:

$$\text{expand}\left(\left(x^3+x^4+x^6+x^9\right)^2\right) \quad x^{18}+2\cdot x^{15}+2\cdot x^{13}+3\cdot x^{12}+2\cdot x^{10}+2\cdot x^{9}+x^{8}+2\cdot x^{7}+x^{6}$$

$$\text{expand}\left(\left(x^3+x^4+x^6+x^9\right)^3\right) \quad x^{27}+3\cdot x^{24}+3\cdot x^{22}+6\cdot x^{21}+6\cdot x^{19}+7\cdot x^{18}+3\cdot x^{17}+9\cdot x^{16}+6\cdot x^{15}+3\cdot x^{14}+6\cdot x^{13}+4\cdot x^{12}+3\cdot x^{11}+3\cdot x^{10}+x^{9}$$

$$\text{expand}\left(\left(x^3+x^4+x^6+x^9\right)^4\right)$$
$$x^{36}+4\cdot x^{33}+4\cdot x^{31}+10\cdot x^{30}+12\cdot x^{28}+16\cdot x^{27}+6\cdot x^{26}+24\cdot x^{25}+19\cdot x^{24}+12\cdot x^{23}+28\cdot x^{22}+20\cdot x^{21}+18\cdot x^{20}+24\cdot x^{19}+14\cdot x^{18}+12\cdot x^{17}+13\cdot x^{16}+8\cdot x^{15}+6\cdot x^{14}+4\cdot x^{13}+x^{12}$$

$$\text{expand}\left(\left(x^3+x^4+x^6+x^9\right)^5\right)$$
$$\cdots 10\cdot x^{35}+50\cdot x^{34}+45\cdot x^{33}+30\cdot x^{32}+80\cdot x^{31}+61\cdot x^{30}+60\cdot x^{29}+95\cdot x^{28}+65\cdot x^{27}+70\cdot x^{26}+85\cdot x^{25}+60\cdot x^{24}+60\cdot x^{23}+55\cdot x^{22}+35\cdot x^{21}+31\cdot x^{20}+25\cdot x^{19}+15\cdot x^{18}+10\cdot x^{17}+5\cdot x^{16}+x^{15}$$

Various ways to get a sum of 20 (or not)

If you want to find the number of ways to roll 20 with *any* number of rolls, you need the coefficient of x^{20} in

$$\sum_{n=0}^{\infty} \left(x^3 + x^4 + x^6 + x^9\right)^n$$

But this is a formal geometric series with common ratio $x^3 + x^4 + x^6 + x^9$, so it sums to

$$\frac{1}{1 - (x^3 + x^4 + x^6 + x^9)}$$

A CAS can now read out the number of ways of getting any sum by generating enough terms of the Taylor series:

$$\text{taylor}\left(\frac{1}{1-\left(x^3+x^4+x^6+x^9\right)}, x, 20, 0\right) \quad 1+x^{3}+x^{4}+2\cdot x^{6}+2\cdot x^{7}+x^{8}+4\cdot x^{9}+5\cdot x^{10}+3\cdot x^{11}+8\cdot x^{12}+12\cdot x^{13}+9\cdot x^{14}+17\cdot x^{15}+27\cdot x^{16}+25\cdot x^{17}+38\cdot x^{18}+61\cdot x^{19}+64\cdot x$$

64 ways to get a sum of 20

Of course, the same idea works with any die with any number of faces. Imagine a die with only two faces—a coin, say—numbered 1 and 2. Then the generating function for the Kerins game is

$$\frac{1}{1 - (x + x^2)}$$

But this is precisely the generating function for the Fibonacci numbers, a fact we'll prove in the next subsection.

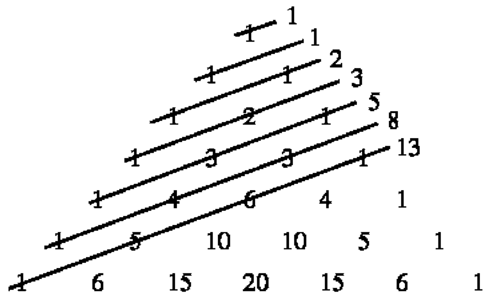
$$\text{taylor}\left(\frac{1}{1-\left(x+x^2\right)}, x, 10, 0\right)$$
$$1+x+2\cdot x^{2}+3\cdot x^{3}+5\cdot x^{4}+8\cdot x^{5}+13\cdot x^{6}+21\cdot x^{7}+34\cdot x^{8}+55\cdot x^{9}+89\cdot x^{10}$$

Fibonacci numbers as coefficients

There's a bonus here. Using the formula for the sum of a formal geometric series, we have

$$\frac{1}{1 - (x + x^2)} = \sum_{n=0}^{\infty} (x + x^2)^n$$

If each binomial on the right-hand side is expanded, it's coefficients are elements of Pascal's triangle. And gathering up all the coefficients of, say, x^6, we express Fibonacci number 13 as a sum of binomial coefficients. This explains the appearance of Fibonacci in Pascal:



Fibonacci as sums of "lazy diagonals" in Pascal.

Fibonacci numbers. Here, we sketch a derivation the generating function for the Fibonacci numbers.

Suppose that $F(n)$ is the nth Fibonacci number, defined by

$$F(n) = \begin{cases} 1 & n = 0 \\ 1 & n = 1 \\ F(n-1) + F(n-2) & n > 1 \end{cases}$$

and let $G(x)$ be defined by

$$G(x) = \sum_{n=0}^{\infty} F(n)x^n = 1+x+2x^2+3x^3+5x^4+8x^5+13x^6+\dots$$

Then, multiplying by x and then x^2, we have

$$\begin{aligned} G(x) &= 1 + x + 2x^2 + 3x^3 + 5x^4 + 8x^5 + 13x^6 + \dots \\ xG(x) &= \quad x + x^2 + 2x^3 + 3x^4 + 5x^5 + 8x^6 + 13x^7 + \dots \\ x^2G(g) &= \quad\quad x^2 + x^3 + 2x^4 + 3x^5 + 5x^6 + 8x^7 + 13x^8 + \dots \end{aligned}$$

Note that these are all calculations with formal power series.

Hence

$$\begin{aligned} G(x) - xG(x) - x^2G(x) &= 1 \quad \text{so,} \\ G(x)(1 - x - x^2) &= 1 \quad \text{and,} \\ g(x) &= \frac{1}{1 - x - x^2} \end{aligned}$$

Probabilistic Number Theory

The problem sets develop two avenues to the calculation of the probability $\mathfrak{P}$ that two integers chosen at random are relatively prime (or equivalently, that a given random integer is square free):

- Using a geometric version of the fact that, for non-

See Problem Set 9.

negative integers k, a, and b, if $\gcd(a, b) = 1$, then $\gcd(ka, kb) = k$, participants find that

$$\mathfrak{P} + \frac{1}{4}\mathfrak{P} + \frac{1}{9}\mathfrak{P} + \frac{1}{16}\mathfrak{P} + \cdots = 1$$

- Ruling out the divisibility of a pair (a, b) by any prime p, leads to

See Problem Set 14.

$$\mathfrak{P} = \frac{1}{\displaystyle\prod_{p \text{ prime}} 1 - \frac{1}{p^2}}$$

Both of these results lead to the same fact:

$$\frac{1}{\mathfrak{P}} = \sum_{n=1}^{\infty} \frac{1}{n^2}$$

Euler found the exact value of the sum on the right side of this equation:

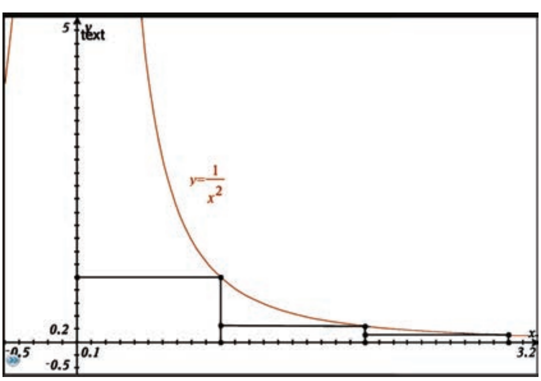
$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{1}$$

This implies that $\mathfrak{P} = \frac{6}{\pi^2} \approx 0.60792710185404$, which agrees with the approximations that participants computed at PCMI.

Euler's first derivation of equation 1 involved expressing $\frac{\sin x}{x}$ as an infinite product, and even he was worried about some of the underlying assumptions (see [3] for the details). Many other derivations of equation 1 have been found over the centuries (including several by Euler himself). Let's look at one of them.

An evaluation of $\sum_{n=1}^{\infty} \frac{1}{n^2}$. First of all, the sum converges, because you can compare it to an integral:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \leqslant 1 + \int_1^{\infty} \frac{1}{x^2}\, dx = 1 + -\frac{1}{x}\bigg|_1^{\infty} = 2$$



$\sum_{n=1}^{\infty} \frac{1}{n^2}$ versus $\int_1^{\infty} \frac{1}{x^2}\, dx$

Our proof comes from a 1972 article in *Mathematics Magazine* by Dan Giesy [2]. Like all derivations of the result, Giesy's uses some trigonometry (addition formulas for sine and cosine) and a little calculus (integration by parts).

Let's start with the trigonometry. We need an identity. Consider the sum:

$$\sin\frac{x}{2}+2\sin\frac{x}{2}\cos x+2\sin\frac{x}{2}\cos 2x+2\sin\frac{x}{2}\cos 3x+\cdots+2\sin\frac{x}{2}\cos nx$$

Using the identity

$$2\sin A\cos B=\sin(A+B)+\sin(A-B)$$

and the fact that $\sin(-x)=-\sin x$, the sum can be rewritten:

$$\sin\frac{x}{2}+\left(\sin\frac{3x}{2}-\sin\frac{x}{2}\right)+\left(\sin\frac{5x}{2}-\sin\frac{3x}{2}\right)+\left(\sin\frac{7x}{2}-\sin\frac{5x}{2}\right)+ \\ \cdots+\left(\sin\left(n+\frac{1}{2}\right)x\right)-\left(\sin\left(n-\frac{1}{2}\right)x\right)$$

This sum telescopes to

$$\sin\left(n+\frac{1}{2}\right)x$$

So, we have

$$\sin\frac{x}{2}+2\sin\frac{x}{2}\cos x+2\sin\frac{x}{2}\cos 2x+\cdots+2\sin\frac{x}{2}\cos nx=\sin(n+\frac{1}{2})x$$

Factoring $2\sin\frac{x}{2}$ out of the left-hand side and dividing, we have

$$\frac{1}{2}+\cos x+\cos 2x+\cos 3x+\cdots+\cos nx=\frac{\sin\left(n+\frac{1}{2}\right)x}{2\sin\frac{x}{2}} \tag{2}$$

Notice that the limit of the right side as $x\to 0$ is $n+\frac{1}{2}$ (use L'Hôpital's rule, for example), so that, if we extend the right side by continuity, this identity even holds at $x=0$.

Next, the calculus. The plan is to take our identity (2), multiply it by x, integrate both sides from 0 to π, and then let $n\to\infty$. That will force the emergence of the sum of the reciprocals of the squares as well as the mysterious π. Here we go:

Multiplying both sides of (2) by x and integrating, we get:

$$\int_0^\pi\frac{x}{2}\,dx+\int_0^\pi x\cos x\,dx+\int_0^\pi x\cos 2x\,dx+\int_0^\pi x\cos 3x\,dx+\cdots+\int_0^\pi x\cos nx\,dx=$$

$$\int_0^\pi \frac{x \sin\left(n+\frac{1}{2}\right)x}{2\sin\frac{x}{2}}\,dx$$

Let's work on the left side first. The first term is

$$\int_0^\pi \frac{x}{2}dx = \frac{\pi^2}{4}$$

Each of the next integrals is of the same form, and it can be evaluated by integration by parts:

$$\begin{aligned}
\int_0^\pi x\cos mx\,dx &= \left.\frac{x\ \sin mx}{m}\right|_0^\pi - \frac{1}{m}\int_0^\pi \sin mx\,dx \\
&= \left.\frac{1}{m^2}\cos mx\right|_0^\pi \\
&= \frac{1}{m^2}\left((-1)^m - 1\right) \\
&= \begin{cases} -\frac{2}{m^2} & \text{if } m \text{ is odd} \\ 0 & \text{if } m \text{ is even} \end{cases}
\end{aligned}$$

So, the sum on the left will not change at even values of n. Hence we can assume n is odd, say $n = 2t+1$. Making this substitution we have

$$\frac{\pi^2}{4} - 2\left(1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\cdots+\frac{1}{(2t+1)^2}\right) = \int_0^\pi \frac{x\sin\left(2t+\frac{3}{2}\right)x}{2\sin\frac{x}{2}}\,dx$$

Finally.... An equation that has both sums of reciprocals of squares (only the odd ones, but at least they are there) and π.

The next step, before we worry about the missing terms, is to let $t \to \infty$. The left side will be $\frac{\pi^2}{4}$ minus twice the sum of the reciprocals of all the odd squares. And it can be shown that the right side goes to 0.

One way to show this is to use a result in analysis that follows from a general theorem due to Riemann. It says that for reasonable functions f (continuous and differentiable on $[0,\pi]$, say), the two integrals

$$\int_0^\pi f(x)\sin tx\,dx \qquad \text{and} \qquad \int_0^\pi f(x)\cos tx\,dx$$

both go to 0 as $t \to \infty$. For example, evaluate the first integral by parts:

$$\int_0^\pi f(x)\sin tx\,dx = -\left.f(x)\frac{\cos tx}{t}\right|_0^\pi + \frac{1}{t}\int_0^\pi f'(x)\cos tx\,dx$$

Each of these terms goes to 0 as $t \to \infty$.

Now,the integral we care about

$$\int_0^\pi \frac{x\sin\left(2t+\frac{3}{2}\right)x}{2\sin\frac{x}{2}}\,dx$$

can be written as

$$\int_0^{\pi} \frac{\frac{x}{2}}{\sin\frac{x}{2}} \sin\left(2t+\frac{3}{2}\right) x \, dx$$

Since $\dfrac{\frac{x}{2}}{\sin\frac{x}{2}}$ is continuous and differentiable on $[0,\pi]$ ($\lim_{x\to 0} \dfrac{\sin x}{x} = 1$), this can transformed into the sum of two integrals like those in Riemann's lemma by applying the addition formula for sine to $\sin\left(2t+\frac{3}{2}\right)x$. So, granting this detail, we see that

$$\frac{\pi^2}{4} - 2\left(1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\dots\right) = 0$$

or

$$\frac{\pi^2}{8} = 1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\dots$$

That's pretty interesting all by itself (you might check it numerically). A little more work gets us Euler's result. Notice that the missing terms are a multiple of the entire sum we want:

$$\frac{1}{2^2}+\frac{1}{4^2}+\frac{1}{6^2}+\dots = \frac{1}{4}\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\dots\right)$$

Hence we have two equaltions:

$$1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\dots = \frac{\pi^2}{8}$$

$$\frac{1}{2^2}+\frac{1}{4^2}+\frac{1}{6^2}+\dots = \frac{1}{4}\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\dots\right)$$

Adding, we get

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\dots = \frac{\pi^2}{8}+\frac{1}{4}\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\dots\right)$$

or

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{8}+\frac{1}{4}\sum_{n=1}^{\infty}\frac{1}{n^2}$$

that is,

$$\frac{3}{4}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{8}$$

so

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$$

Note that this is the reciprocal of the probability that a random pair is relatively prime, the number we are calling $\mathfrak{P}$. So, The answer to our question is:

Theorem 4.2

The probability that a pair or randomly chosen integers is relatively prime is $\frac{6}{\pi^2}$.

The Harmonic Series

The problem sets also develop the idea that the harmonic series diverges, and this can be used to prove that there are infinitely many primes via the product:

$$\prod_{p \text{ prime}} \frac{1}{1-\frac{1}{p}} \text{“}=\text{”} \sum_{n=1}^{\infty} \frac{1}{n}$$

Ken Levasseur of UMASS Lowell pointed us to a simple proof that the harmonic series diverges: If $\sum_{n=1}^{\infty} \frac{1}{n}$ converged to some finite number s, we'd have

$$\begin{aligned} s &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \dots \\ &> \frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \dots \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots = s \end{aligned}$$

Bibliography

[1] Education Development Center. (2009). *CME Project Precalculus*. Boston, MA: Pearson.

[2] Giesy, D.P. (1972, March). Still another elementary proof that $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$. *Mathematics Magazine*, *45*(3), 148–149.

[3] William Dunham. (1999). *Euler, master of us all*. Washington, DC: Mathematical Association of America.