COINCIDENCE IN RUNS AND CLUSTERS

Milo Schield, W. M. Keck Statistical Literacy Project

Abstract: Coincidence may be the most common manifestation of chance in everyday life. Chance-based coincidences are a necessary result of the Law of Very Large Numbers: the unlikely is expected (the "impossible" becomes almost certain) given enough tries. This paper examines: runs in flipped coins, one and two-dimensional clusters of rare events and the birthday problem using Excel spreadsheets. This paper has three goals: 1) to show that unlikely coincidences are much more common than expected, 2) to show the ambiguity in the question "What is the chance of that?" and 3) to show that as a rule-of thumb a run with one chance in N is generally found in a string of N tries. This rule of thumb is simple, memorable and useful since it is within 10% of the expected value for large N Conjecture: if the chance of run of length k is p^k, then that length run is expected in N tries when $N = (1/p)^{(k+1)}$.
Keywords: statistical literacy, Excel, visual display

1. STATISTICAL LITERACY

No matter how one defines statistical literacy, it must include randomness or chance. The two most common references to chance in the everyday media involve the margin of error and coincidences.

Coincidences are newsworthy. Raymond and Schield (2008) analyzed 273 statistics-based news stories. Of these, 10% included the phrase "unlikely due to chance" whereas 5% mentioned "confidence level" and only 3% mentioned "statistically significant".

The goal of this paper is to investigate runs and clusters, and show why coincidences are expected.

2. COINCIDENCES

Coincidences are event-conjunctions that are unlikely and memorable. Life is filled with unlikely events. Recall the first name of the last stranger you talked to. Unlikely? Yes. Memorable? Not likely.

Now recall an unlikely conjunction that was memorable for you. Calling someone just as they were about to call you (or vice versa). Meeting someone you know in a distant or unlikely place. Finding something lost long ago, just before you were going to buy a replacement. Having a dream about something before it happened.

John Allen Paulos, author of *Innumeracy* stated that the most incredible coincidence of all might be the absence of all coincidence. Auditors sometimes use the lack of coincidence as evidence of fraud.

"Coincidences fascinate us; they seem to compel a search for their significance." Paulos (1989). "My favorite is this little known fact: In Psalm 46 of the King James Bible, published in the year that Shakespeare turned 46, the 46th word is "shake" and the 46th word from the end is "spear." (More remarkable than this coincidence is that someone should have noted this!)" Myers (2002).

"One famous coincidence is that John Adams and Thomas Jefferson, two men who shaped the Declaration of Independence, both died on July 4, 1826, the fiftieth anniversary of the signing of that historic document." (Neimark, 2004)

Coincidences that are extremely unlikely due to chance are argued to be due to something else. So how likely are coincidences?

The short answer: "Coincidences are more likely than you ever imagined." The longer answer: "Coincidences are expected given enough tries." The goal of this paper is to illustrate these two points.

Coincidences are a matter of a statistical law: the law of Very-Large Numbers: the "impossible" is almost certain given enough tries. As described by Diaconis and Mostellar (1989), "With a large enough sample, any outrageous thing is apt to happen."

To illustrate the omnipresence of coincidences, this paper begins with runs in flipping a fair coin, examines patterns or clusters of rare events in one and two-dimensional spaces and reviews the classic "birthday" problem. Excel spreadsheets generate the data in this paper.

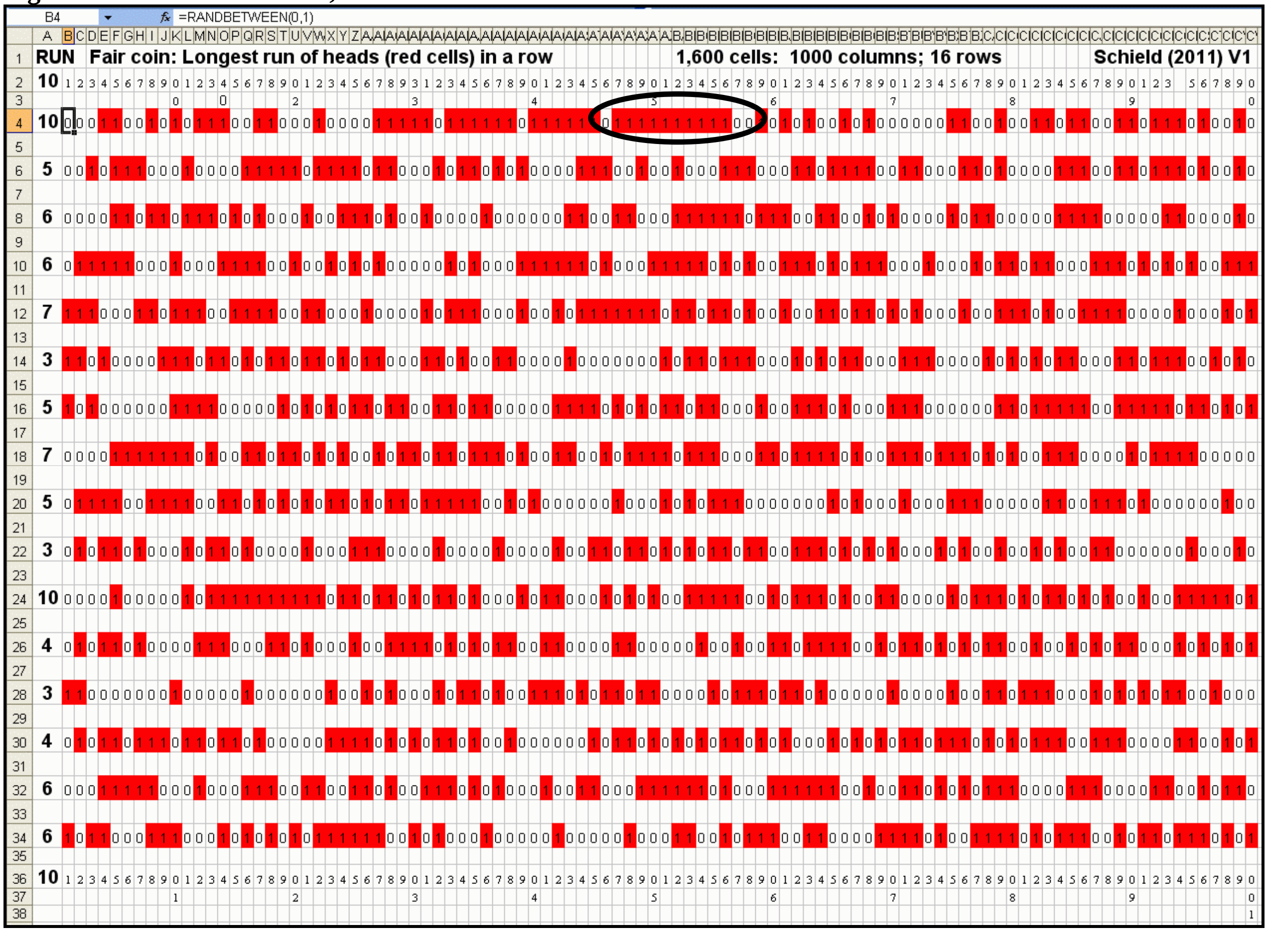
3. COINCIDENCE: RUNS IN FLIPING COINS

Consider seeing a *run* in flipping a fair coin. A "run" identifies a group activity as in a "run" on a bank where depositors withdraw their money in large numbers or when a group of people compete to be first in a 10-K "run". In statistics, a run involves repetitions of a rare outcome. For coins, a run is series of heads (or tails).

Runs in flipping coins are a bit more complex than expected. See Appendix A for technical details.

Figure 1 shows the results of flipping 16 rows of 100 coins each. It shows if the result is a head (a red color), tabulates the length of the longest run in each row, and then show the maximum length of all the runs in that spreadsheet.

Figure 1: Runs of Heads in 1,600 cells



The longest run for each row is shown at the left in column A. The longest run of all is shown at the top left corner.

Notice the run of ten heads circled in the first row of Figure 1. *What is the chance of that?*

A run of 10 heads seems very unlikely. Yet runs of at least ten heads occur more often than not when the data if refreshed (press F9). How is this possible?

The short answer: the question is ambiguous. "What is the chance of that?" has two interpretations:

1. What is the chance that on the *next* 10 tries, heads will result in *exactly* the cells circled?
2. What is the chance that ten adjacent cells will all have heads -- *somewhere* in the 1,600 cells?

The answer to the first question is one chance in 1,024: $P = (1/2)^{10}$. The answer to the second depends on what length longest run is expected in N tries.

The expected length of the longest run is not determined by the binomial distribution. But if there are T trials (where T = 1/P), of 10 flips each, then one run is expected: on average. If P is one chance in 1024, then in 1,024 trials of 10 coins each, one set of 10 heads is expected. It can be shown that in this case the chance of 10 or more heads is at least 50% (is more likely than not). See Appendix B and Appendix C. When Np is an integer, Lord (2010) showed that the mean is also the median and the mode.

But the Binomial distribution does not give any information about the distribution of the longest run in a string of size N. But knowing that this coincidence (a run of k heads) is expected in T sets of size k is a useful first step that matches with our intuitions.

Consider a simpler case of a run of 3 heads: one chance in eight. A run of three heads is expected in T trials of 3 coins each where T = 8 since T*P = 1. This would involve flipping 24 coins.

These eight sets of 3 coins each can be mapped onto a series of 10 coin flips as shown in Figure 2.

In general, T trials of K coins each can be mapped into N flips when N = T + (k-1) and $T = 2^k$. For runs of length k, consider $N = (1/p)^k + (k-1)$.

Figure 2: Compressing 24 flips (8 sets of 3 each) into 10 flips in a row

A	B	C	D	E	F	G	H	I	J	K	L
Run of 3 heads generally found in 10 flips of a fair coin.									Schield (2012) V1		
Coin	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	
RUN											
3	1	1	1	0	1	0	1	1	0	1	
3	1	1	1								Set #1
2		1	1	0							Set #2
1			1	0	1						Set #3
1				0	1	0					Set #4
1					1	0	1				Set #5
2						0	1	1			Set #6
2							1	1	0		Set #7
1								1	0	1	Set #8
	Distribution of longest run of heads in a set of 3										
	Longest Run	0	1	2	3						
	Expect #	1	4	2	1						
	Pctg of 8	12.5%	50.0%	25.0%	12.5%						
		TTT	HTT, THT,	HHT, THH	HHH						
			TTH, HTH.								

When the event is a run of two heads (chance of one in four) with N = 5, the chance of that run or longer is greater than 50% and the mean length is 1.94.

When the event is a run of three heads (one chance in eight) with N = 10, the chance of that run or longer is greater than 50% and the mean length is 2.80.

Appendix E thru Appendix J support the claim that the longest run generally found in series of N flips of a fair coin is given by Log(N) base 2 for N >> k.

Table 1 shows the number of flips generally needed given the length of the longest run.

Table 1 Number of flips needed by longest run length

Length	N		Length	N
Two	4		Seven	128
Three	8		Eight	256
Four	16		Nine	512
Five	32		Ten	1,024
Six	64		Eleven	2,048

With this background, we can return to the horizontal runs in Figure 1. In a row of 100 cells, we generally find a maximum run length of six or seven. In this figure, there are 16 rows. A simple way to handle this situation is to treat the 16 rows as being linked, so there is a single string of 1,600 tries. In that case, we generally find a longest run length of 10.

This is what is observed in Figure 1.

Alternatively, when the run of interest is one chance in N, then one needs N cells so that one such run is generally found. This simple rule of thumb is intuitive, memorable and visually convincing.

If the chance of success on the next try is p, the chance of a run of k or more successes is generally more likely than not in N tries where $N = (1/p)^k$.

Note the relentless use of “generally.” This is a rule of thumb, not a mathematical theorem. .

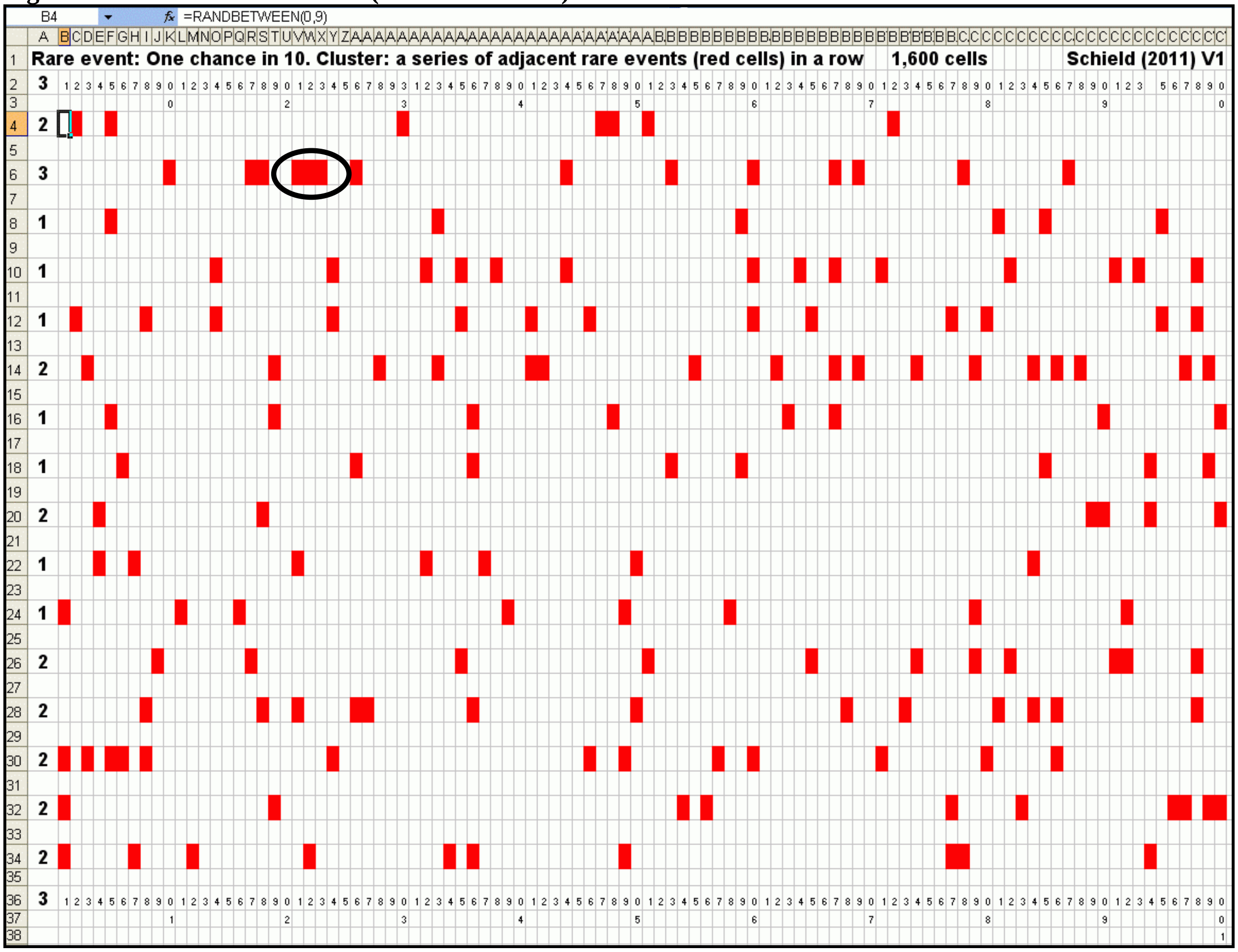
This run of 10 in Figure 1 is unexpected for the non-statistician who doesn’t see the ambiguity in the question and who fails to see the large number of potential groups of size 10 within these 1,600 cells. This run of 10 heads in more than a thousand tries is not unexpected by anyone who is statistically-literate.

Recall again the ambiguity in the question: How likely (unlikely) is that? This ambiguity explains why coincidences can be both rare and common. Appendix D examines the distribution of run lengths.

The real key to these very long runs is overlap. To expect a run of 10 heads, it would take 1,024 tries of 10 coins each. Instead this run of 10 heads is generally found in a row with 1,024 tries. Overlap within a row gives a factor of 10 reduction.

Coins are easy to understand, but powers of two are not. Consider events having one chance in 10.

Figure 3: Cluster of Rare Events (one chance in 10) in a Row



4. STRAIGHT-LINE CLUSTERS

Clusters describe a group of rare events that are connected. In this case, a cluster occurs when the rare events are connected by touching horizontally.

This can be simulated in Excel by using a modified version of the runs spreadsheet where the chance of a rare event is one chance in 10. This is simulated using the Excel function RandBetween (0,9). Getting a nine is considered a rare event: one chance in ten.

In this spreadsheet each cell has a one-digit value between zero and nine. Those values involving a nine are highlighted in red. Note that the rare events (the red cells) are much less common – as expected. First consider straight-line clusters that are horizontal.

Now consider the spreadsheet results shown in Figure 3. The number in the left-hand column (column A) is the maximum run length in that row. The number in the lower-left and upper-left corners is the maximum run length for all the rows. Notice that the maximum horizontal run-length is one for some of the rows, two for a few and three for a very few.

Now consider the three-event cluster circled in the second row. What is the chance of that?

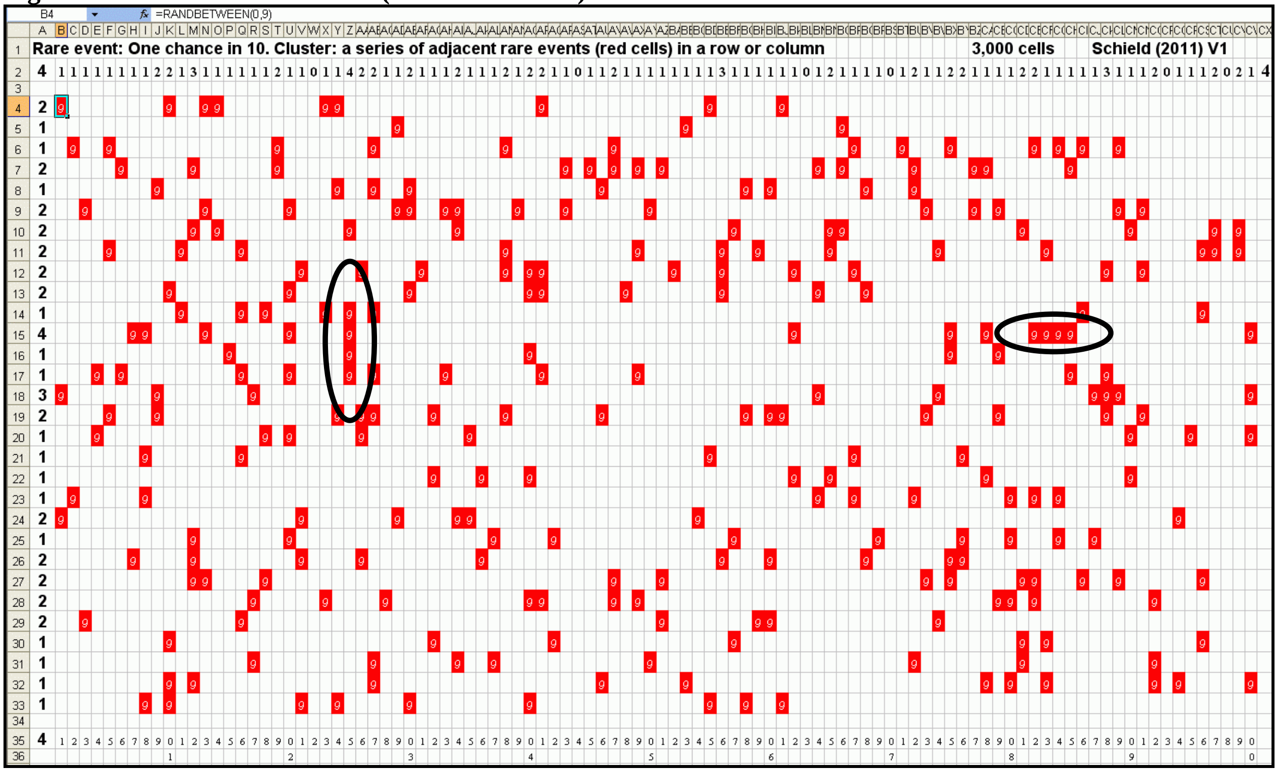
An unwitting student might say: one chance in one-thousand since each rare event has one chance in ten. This is a correct answer – in a sense. As before, this question is ambiguous. You can see this by refreshing the data (press F9) and seeing that having at least one three-event cluster is common. You are getting a thousand-year flood every year.

In fact you may be getting a four-event cluster (one chance in ten thousand). How is this possible?

As in the case of the coins, the real problem is the ambiguity of the question. So what is the probability of the three-event cluster circled in the first row?

1. What is the chance that three rare events will be adjacent – in the *specified location*: the circled cells? One chance in 1,000: $P = 10^{-3}$.
2. What is the chance that three rare events will be adjacent -- *somewhere* in the 1,600 cells?

Figure 4: Cluster of Rare Events (one chance in 10) in a Row or Column



The answer to the second question depends on how many opportunities there are to get three rare events in adjacent cells. Suppose there were 1,000 such opportunities.

What is the size of the largest cluster one would generally find? Since 10^3 = 1,000, one generally finds at least one three-event cluster. We generally find a thousand-year flood every year somewhere on earth given a thousand independently floodable places. This event (or higher) is more likely than not.

Table 2 Distribution of Row Clusters (N=1,000)

RUN	Probability: P	Chance: P	k = N*P
One	One in 10	(1/10)^1	100
Two	One in 100	(1/10)^2	10
Three	One in 1,000	(1/10)^3	One
Four	One in 10,000	(1/10)^4	1/10

Clusters of rare events are much more common than expected provided one is aware of the size of the region in which such clusters could appear. Unfortunately, the resulting cluster is seen while the region in which it could have appeared in unseen or unnoticed.

Now consider straight-line clusters in a single column (vertical). This would have exactly the same chance as getting a cluster in these rows. Including both rows and columns doubles the number of possible ways to get a straight-line cluster of adjacent rare-event cells in a given area. See Figure 4.

A straight-line cluster of three rare events (p=1/10 so P=1/1,000) horizontally or vertically is generally found in some 500 cells.

Here the key is overlap – but in a different sense. A given cell can form a coincidence by row – or by column. This is what gives the factor of two.

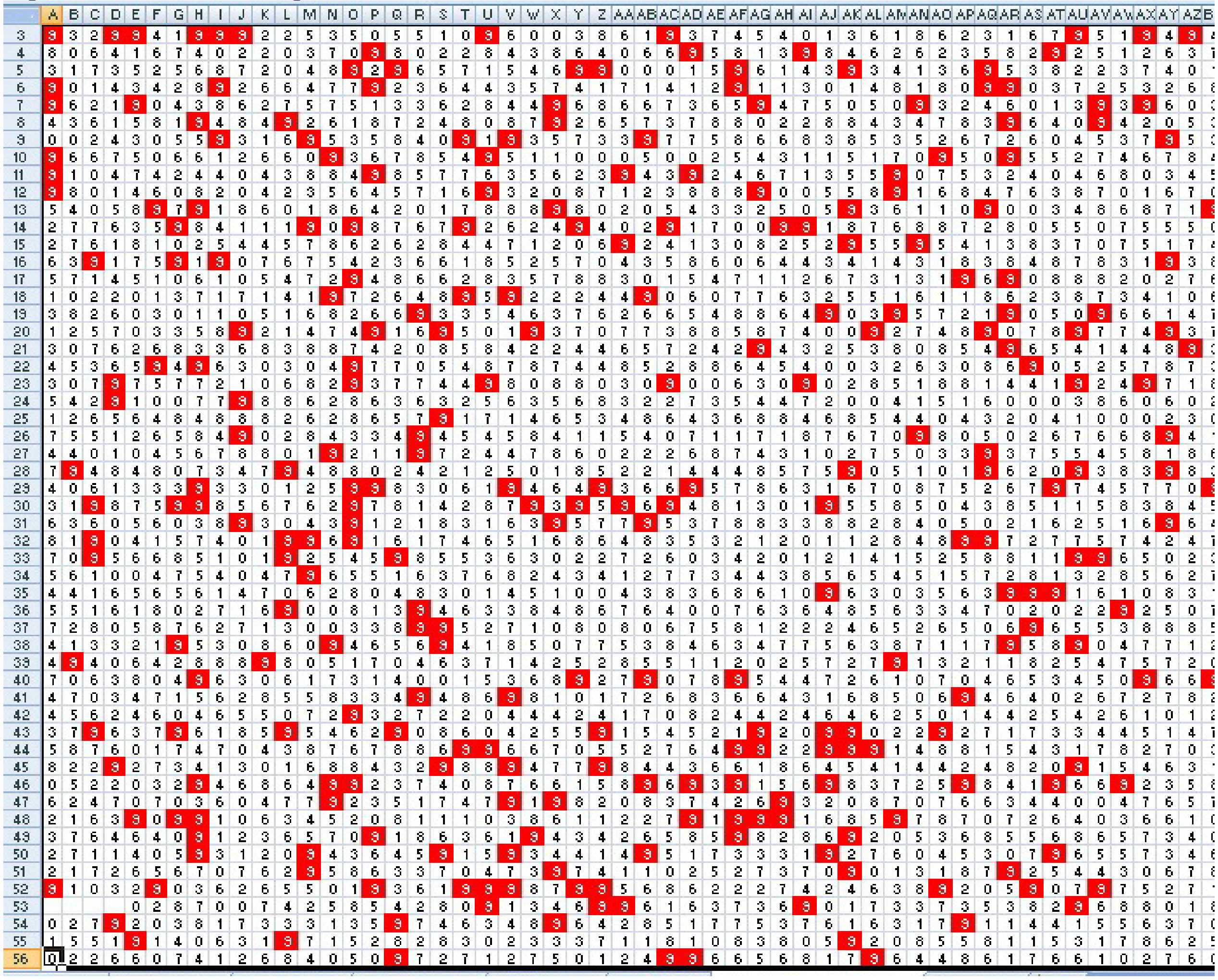
Now suppose we allowed clusters whenever two cells touch. This would allow diagonal connections. For a given cell there are two horizontally (H) adjacent cells and two vertically (V) adjacent cells, but there are four diagonally (D) adjacent cells. Allowing a straight line cluster diagonally quadruples the number of ways to get a straight-line cluster of adjacent cells.

Table 3 Trials Needed (p=1/10) by Run Length

Length	N (H)	N (H or V)	N (H, V or D)
Two	100	50	13
Three	1,000	500	125
Four	10,000	5,000	1,250
Five	100,000	50,000	12,500
Six	1million	500,000	125,000
Seven	10 million	5 million	1.25 million

The more ways there are to connect things, the physical fewer trials required for a logical conjunction.

Figure 5: Distribution of "grains of rice"



5. RICE CLUSTERS IN 2 DIMENSIONS

Suppose that we did not require that the rare events be in a straight line. This should increase the possible ways to generate a cluster and decrease the number of trials needed to get a cluster of a given size.

In this spreadsheet each cell has a randomly-assigned one-digit value between zero and nine. Only those values involving a nine are highlighted in red. The chance a given cell will be red is one chance in ten.

There does not appear to be any formula to generate the number of ways that k tiles can touch each other starting with a given cell where touching includes points as well as edges. Simple overlap compression doesn't capture the ways cells can form patterns. Appendix K gives an upper estimate of this number.

We don't have any idea of how far over the true value the upper-limit is. Eliminating the last tile in the branching gives $8x[7^{(k-3)}]$ combinations. Table 4 gives these estimates of the number of combinations possible with k tiles as a function of k.

Table 4 Estimated # of Combinations for k Tiles

k	$8*[7^{(k-3)}]$		k	$8*[7^{(k-3)}]$
2	1		7	19,208
3	8		8	134,456
4	56		9	941,192
5	392		10	6,588,344
6	2,744		11	46,118,408

How many ways are there to get a cluster of nine events in a spreadsheet with 3,000 cells? First eliminate a band of 9 cells around a horizontal and vertical side to compensate for edge effects. If there are one million ways to form a cluster of nine events, then there are more than one billion (10^9) ways to get a cluster of nine cells touching with 3,000 cells. This means at least one cluster of nine or more touching cells ($P=10^{-9}$) is generally found. .Students enjoy working with the "grains of rice" since it generates unusual patterns. They quickly realize that they are getting extremely-unlikely coincidences every time. It's like a billion-year flood every year. See Figure 5.

6. COINCIDENCE: BIRTHDAY PROBLEM

The classic example of coincidence is the "birthday problem." The chance of a match in birthdays (month and day) seems remote: 1 chance in 365.

Note the ambiguity in "match". Here a match means a shared birthday (same month and day) for two members of a group. It does not mean an exact match with a pre-specified birthday (month and day).

Richard von Mises proved that a shared match was expected in a group of at least 28 people. See Schield and Burnham (2008). Von Mises noted the sheer number of connections and the fact that the matched subject were specified after the fact (ex post) – not in advance (ex ante). Since the matched subjects were not specified in advance, the probability of a match was the probability that any two subjects would have matching birth dates. This probability depends on the chance of a given birth data (1/365) and the number of possible pairs or subjects.

The number of pairs (the number of permutations) is easily calculated. For a group of N subjects, the first subject has N-1 connections; the second has N-2 additional connections, etc. So the total number of order-dependent connections is given by:

Eq. 1 # permutations N things 2 at a time: N(N-1)

To eliminate order (to calculate the number of combinations), we must divide by the two places to put the second choice.

Eq. 2: # combinations N things 2 at a time: N(N-1)/2

For N = 28, the number of connection is 378: 27*14.

Table 5 shows the number of connections for various size groups (N).

Table 5 Combinations of Two in Groups of Size N

N	Connections	N	Connections
2	1	26	325
3	3	27	351
5	10	28	378
10	45	30	435
25	300	50	1,225

An event with probability P is *expected* in 1/P tries. An event with probability 1/365 is *expected* in 365 tries. Note that the probability of a match is 1/365 – not $(1/365)^2$. This is because a shared "match" is different than the possibility of an exact match with a pre-specified month and day.

Lesser (1999) noted that the exact solution where the events are not independent is quite close to the approximate solution where the events are independent. Henceforth, the trials will be treated as independent.

Consider an exact match. It takes at least 253 random picks to have at least a 50% chance of a match with a pre-specified date. 1- [(364/365)^253] = 0.51

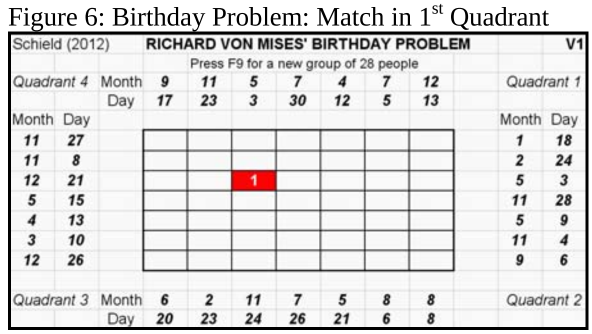
The expected value is not necessarily what happens most of the time; it need not be the most common outcome (the mode). It is what is expected on average – in the long run. The chance of two heads in two flips of a fair coin is 1 in 4. In four flips of a pair of coins, we expect one pair of heads

As mentioned before, an expectation can also be expressed as a probability statement. An event with probability P is more likely than not to occur at least once in 1/P tries. (Appendix B) So a matching birthday is expected (at least one or more is more likely than not) in a group of at least 28 people.

Although the birthday problem is an unexpected coincidence, students may have a hard time seeing all the possible connections that are responsible for an expected match in a group of only 28 people. Consider the following Excel spreadsheet:

In Figure 6, 28 people are arranged around a table with seven on each side. A match between individuals on adjacent sides is indicated by a cell in the center colored red along with a number indicating the quadrant involved.

Figure 6: Birthday Problem: Match in 1st Quadrant

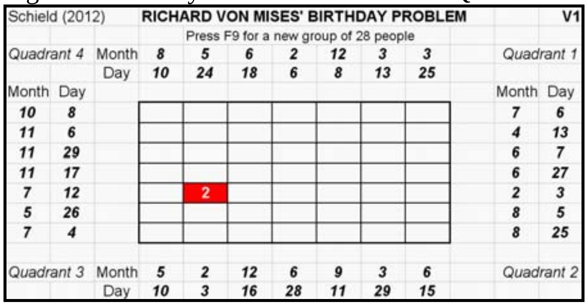


The table is divided into four quadrants. The upper-right quadrant is designated as quadrant 1. The lower-right quadrant is designated as quadrant 2. The lower-left quadrant is designated as quadrant 3. The upper-left quadrant is designated as quadrant 4.

A match involving people in the upper-right quadrant (top and right) is shown by the number "1" in a red-filled cell. There are 49 possible pairs in quadrant 1 between the seven on top and the seven on the right.

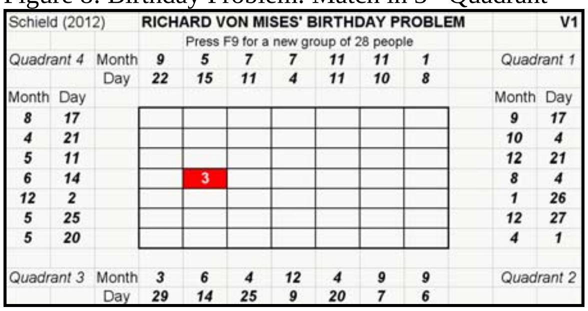
In Figure 7, there is a match between a person on the right and a person on the bottom. This match in quadrant 2 is indicated by the number two in the red-filled cell. There are 49 possible pairs in this second quadrant involving the seven people on the right with the seven people on the bottom.

Figure 7: Birthday Problem: Match in 2nd Quadrant



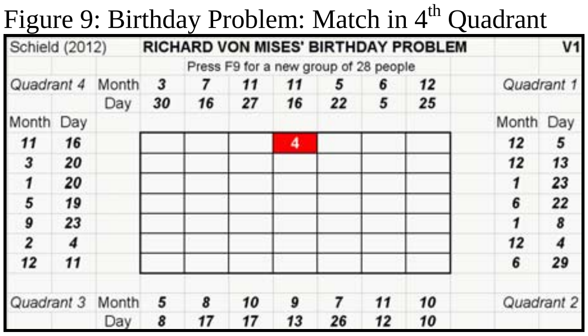
In Figure 8, there is a match between a person on the left and a person on the bottom. This match in quadrant 3 is indicated by the number three in the red-filled cell. There are 49 possible pairs in this third quadrant involving the seven people on the left with the seven people on the bottom.

Figure 8: Birthday Problem: Match in 3rd Quadrant



In Figure 9, there is a match between a person on the left and a person on the top. This match in quadrant 4 is indicated by the number four in the red-filled cell. There are 49 possible pairs in this fourth quadrant involving the seven people on the left with the seven people on the top.

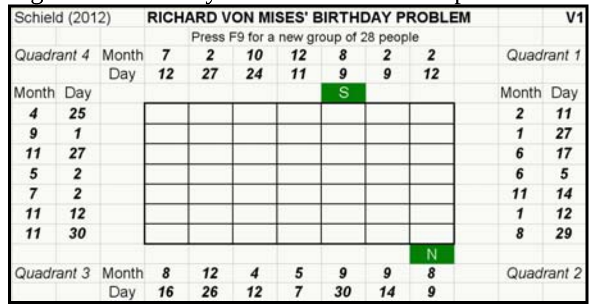
Figure 9: Birthday Problem: Match in 4th Quadrant



At this point we have highlighted 196 possible pairings (4 * 49) between these 28 people. But to make a match more likely than not, we need at least 360 possible pairings. Now consider pairings between individuals on the top and bottom:

Figure 10 illustrates a match between non-adjacent sides: specifically between top (N) and bottom (S).

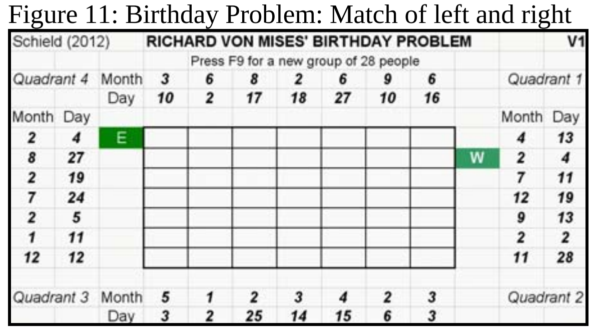
Figure 10: Birthday Problem: Match of top & bottom



A match between two non-adjacent sides is indicated by green-filled cells. The letter indicates the position of the matching cell. The cell on the bottom (South) marked "N" matches with a cell on the top (North) marked "S". Again, we have 49 possible pairs between top and bottom.

Figure 11 shows matches between individuals on opposite sides:

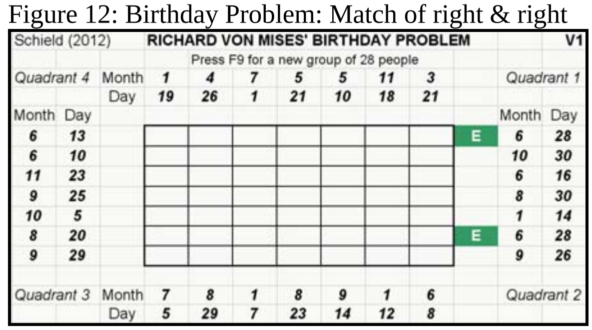
Figure 11: Birthday Problem: Match of left and right



The cell on the left (West) marked "E" matches with a cell on the right (East) marked "W". Again, we have 49 possible pairs between top and bottom.

At this point we have 294 possible pairings: 6 * 49. But we need more to reach our goal. Now consider matches within a given side:

Figure 12: Birthday Problem: Match of right & right



In Figure 12, the cell on the right (East) marked "E" matches with another cell on the right (East) marked "E". Here we have only 21 possible pairings: 6+5+4+3+2+1.

When we include the 84 internal pairing with all four sides we have 378 possible pairs.

Table 6 Connections in a Group of 28 People

Pairs	Location of Connection
196	Quadrants 1 – 4. 49 pairs per quadrant
98	Top to bottom (49) and side-to-side (49)
84	Within each of four sides (21 each)
378	TOTAL

Since this total exceeds 365, a birthday match (a prior probability of one in 365) is *expected* in a group of 28 people.

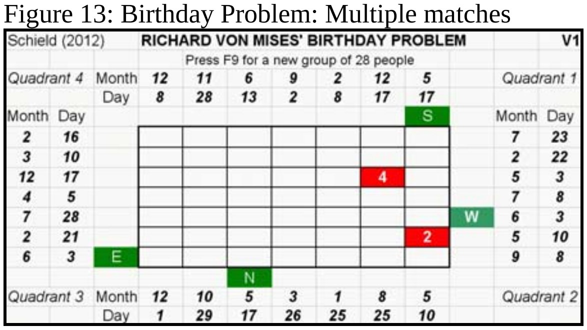
Once again, we don't see all the possibilities. We focus on the observed outcome and ask "What is the chance of that?" By now you should recognize this question as being ambiguous?

It can mean, "What is the chance of specifying exactly two previously identified people in a group as having a matching birthday?" This is the one chance in 365. There is only one way for this match to occur.

It can mean "What is the chance of at least one birthday match among any of the people in a group?" In a group of 28 people, the answer is "more likely than not": greater than 50%.

The actual case is even more complex. There can be multiple matches. Here are some examples:

Figure 13: Birthday Problem: Multiple matches



It is false to say that having a [single] match [having exactly what is expected] is more likely than not when N = 1/P. See Appendix C. The proper statement is that at least one match is more likely than not. See also Schield (2005).

7. COINCIDENCE AND SAMPLING ERROR

At this point it seems like coincidence and sampling error are at opposite ends of a spectrum. Coincidence requires large groups (coincidences increase as the group get bigger); sampling error requires small groups (sampling error increases as the group gets smaller.

But there is a connection. Sampling error is a distribution. All too often we focus on the 95% margin of error. But the sampling distribution can be used to talk about the chance of extremes (coincidences) as well as to talk about the chance of a given margin of error. Consider the case of a cancer cluster.

8. CANCER CLUSTER

In their great book, *The Numbers Game*, authors Blastland and Dilnot tell the story of Wishaw: a small village in England. On bonfire night in 2003, someone tore down the cell-phone tower in their town. No one confessed; no one was ever charged with the crime. It may have been related to the fact that "among the 20 households within 500 yards of the tower, there had been nine cases of cancer."

What is the chance of that? What is the chance that 20 families would have nine cancer victims? It seems like this is too unlikely to be due to chance. If so, this is no coincidence; there must be a "causal explanation" where the quotes indicate a special case in which the causes are readily observable.

But how unlikely is this event? To answer this, we need to sort out some messy issues.

First, what constitutes being a cancer victim? This could range from being a terminal cancer patient to having a benign skin cancer. Second, how many people are there in these 20 families? 20 people or a hundred? Third, how many of these cancer victims were diagnosed with cancer prior to the arrival of the cell-phone tower? Fourth, over what length of time are these cancers tabulated: a year or a decade? Fifth, how many have a cancer that is closely associated with a well-known cause: smoking and lung cancer, sun exposure and skin-cancer, etc.

But, to make a point, assume the cancers were moderate to serious, that there are 50 people in these families, that all the cancers were diagnosed after the tower was installed and none of the cancers had a known cause. Furthermore, assume there is one chance in ten of being a cancer victim for adults.

There are two ways to approach this problem. One uses the sampling distribution; the other using a combinatorial approach.

Consider using the sampling distribution. Suppose that the observed rate (20%) is twice the background rate (10%). How likely is a rate of at least 20% in samples of size 50 when sampling from a population with a rate of 10%? The 95% margin of error is given by 1.96*Sqrt(0.1*0.9/50) = 0.0832 = 8.32%.

If Z = 2.4, then the upper-end of the confidence interval exceeds 20%. The probability of this is eight chances in 1,000. So in 122 groups of size 50, one would expect to find double the background rate.

This coincidence is all but expected – somewhere in a country the size of England.

9. BIG DATA

It might appear that as the size of the dataset increases, the chance of a coincidence should decrease. Consider flipping two coins. The chance of a match on the first pair of flips is one in two. The more times this pair of coins is flipped, the less likely that the resulting series will match.

Eq. 3: P(match of length k|2 coins) = p^k = (1/2)^k

But what if the big data involved an increase in the number of coins being flipped? Now coincidence should increase. The chance of a match on the 1st try among trials involving n different coins is given by:

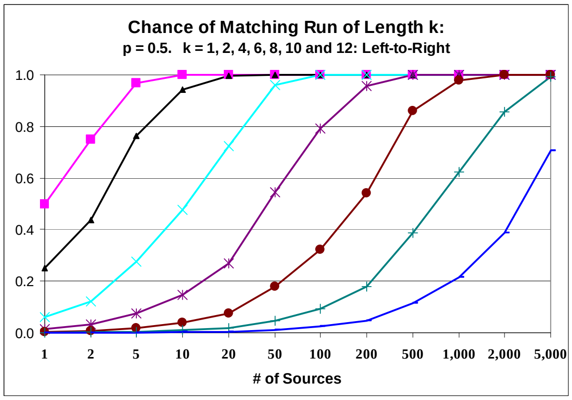
Eq. 4: P(match on 1st try) = 1 – [P(no_match)]^n
P(match on 1st try) = 1 – [1 - P(match)]^n

Inserting the results for a match of length k gives:

Eq. 5: P(match length k | n coins) = 1 – (1-p^k)^n

As k increases, the chance of a match decreases, but as n increases, the chance of a match increases.

Figure 14: Runs: Chance of matching run somewhere



To see the relationship between p, k and n, solve for P(match of length k given n coins) = ½.

Eq. 6: ½ = (1-p^k)^n

As noted by Alex Schield (private communication), this is readily reduced by taking logs which gives (where N = p^-k):

Eq. 7: Log(½) = n * log(1-1/N)

Taking the first term of the Maclaurin series for LN(1-x) and using LN(1/2) = -.693 gives:

Eq. 8: 0.693 ≅ n / N or n ≅ 0.7*N

This approximation is nicely memorable. Given a run with probability 1/N, finding a matching run in 'n' other series is more likely than not if n > N.

This derivation does not say anything about the size of a longest run of length k in K flips of n coins.

10. CAUSATION

When one can show that an unexpected event is expected assuming pure chance, does this prove it is a statistical coincidence? Certainly not.

One cannot prove an observed event is pure coincidence after assuming randomness any more than one can prove the null hypothesis is true after assuming that it is true.

So who has the burden of proof after an event has been shown to be expected assuming pure chance?

- One group says that those who assert causation have the burden. This test is more like criminal law where the presumption is coincidence and strong arguments (beyond reasonable doubt) must be mustered to overcome that presumption.
- The other group says that neither group has the burden of proof. The test is more like civil law where the preponderance of evidence must be used.
- A third group says it depends on the context. If we are doing pure science and want to minimize Type 1 error, then the criminal law approach is best. If we need to make a decision and the stakes are high (life vs. death), then a civil law approach is best.

Sorting this out will be left for later.

11. EXPLANATIONS

It seems that we underestimate the chance of coincidences. Why is this? Are we necessarily error-prone? Do we have minds that cannot handle these mathematically-sophisticated events?

What are some explanations for under-estimating the chance of a coincidence?

1. Ex Ante vs. Ex Post: When someone asks, "What is the chance of that?" are they asking ex ante (before this happened) or ex post (after this happened)? Consider shooting targets with a bulls-eye: Ex ante is posting the target first and then shooting. Ex-post is shooting first and then centering the target on the center of your shooting (much easier).
2. Change in context (here vs. somewhere): When someone points out a coincidence and asks "What is the chance of this?" the 'this' can mean this outcome as just this location (here) or it can mean this outcome at any location (somewhere).
3. Combinations vs. permutations: Consider flipping two coins where there was a head on the first try and a tail on the second. . Asking "what

is the chance of that?" is ambiguous. If one means "exactly this" you are talking about permutations: order matters. If one means one head on either the first or second try, you are talking about combinations: order does not matter.

4. Physical vs. logical. Consider flipping three coins. A set of three heads is "expected" in flipping eight sets of three: 24 flips. But with overlap into a single row, these eight sets can be mapped to 10 physical flips: 123, 234, 345, etc. For a run of size k heads, the $k*2^k$ individual flips are mapped into $2^k + (k-1)$ flips: roughly a k-fold reduction. Generalizing, for a run of size k with a probability p of a success per individual event, the $k*(1/p)^k$ logical trials are mapped on to a $(1/p)^k + (k-1)$ ordered tries in a row. For a triplet with one chance in 10 per try, one would expect one triplet in a thousand sets of three to form a coincidence. These 3,000 tries can be mapped into 1,002 physical trials: roughly a three-fold reduction. The longer the run, the greater the reduction.

5. Psychological explanations. "We are hard-wired to overreact to coincidences," says Persi Diaconis. (Selkin, 2002)

6. Big Data: As the supply of data increases, the number of beguiling coincidences also increases. This increased number may over-whelm even the most skeptical observers. As Professor Robbins (Harvard) noted, "It's well known that if you take a lot of random noise, you can find chance patterns in it, and the Net makes it easier to collect random noise."

7. Not-very-relevant question. Consider a front-page story in the New York Times on a "1in 17 trillion" long shot, speaking of a woman who won the New Jersey lottery twice. Diaconis and Mosteller (1989) took a radically different approach when they said "The 1 in 17 trillion number is the correct answer to a not-very-relevant question. If you buy one ticket for exactly two New Jersey state lotteries, this is the chance both would be winners." They noted, this very specific situation is seldom what has happened. In this case, the woman had bought multiple tickets repeatedly.

Diaconis and Mosteller summed it up this way: "Once we set aside coincidences having apparent causes, four principles account for large numbers of remaining coincidences: hidden cause; psychology, including memory and perception; multiplicity of endpoints, including the counting of "close" or nearly alike events as if they were identical; and the law of truly large numbers, which says that when enormous numbers of events and people and their interactions cumulate over time, almost any outrageous event is bound to occur. These sources account for much of the force of synchronicity."

This paper argues that even when the law of very large numbers is known, the large number may still be unseen – even when visible. Consider this visual explanation.

12. VISUAL EXPLANATIONS

Explanations of coincidence can be quite sophisticated. They may satisfy some but not others. Consider a visual approach to see how what is seen – but ignored (unseen) – matters in explaining coincidences.

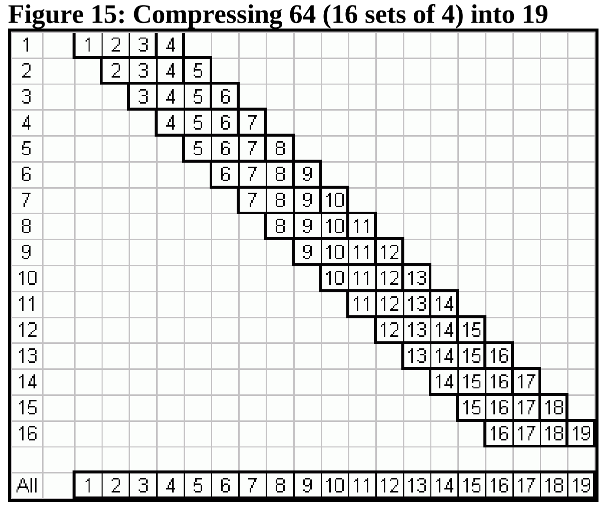
Review the discussion on how a run of three heads is expected in 8 sets of three coins each – and how it is "expected" in a series of 10 coin flips.

Suppose we flip 10 coins? What is the chance that we will get a run of three heads?

The chance of getting a run of three heads somewhere in 10 flips is the same as the chance that flipping eight triplets separately will give a triplet of all heads.

Compressing 24 into 10 may not seem like much.

Figure 15: Compressing 64 (16 sets of 4) into 19



Consider a run of four heads. What is the chance of getting four heads in the next four flips of a fair coin? This is one chance in 16: 2^4. If there were 16 sets of four coins each and we flipped them all one time, we would expect one of the 16 sets to have all heads. That involves flipping 64 coins since 64 = 4*16.

If instead of specifying exactly where or when the run is to occur, we allow the run to occur anywhere in a series of flips, how few coins do we have to flip?

Figure 15 shows this visually. Consider the pattern that is emerging. The chance of getting k heads on the next k flips of a fair coin is $(1/2)^k$. Suppose we had N set with k coins each where $N = 2^k$. If we flipped these N sets of k coins each, we would expect one of those N sets to have all heads.

But these N sets of k coins each can be mapped or compressed into N + (k-2) adjacent cells. So what started as N*k flips ($k*2^k$) has now been reduced to N+k-2 cells ($2^k + k – 2$).

For runs, the ratio of uncompressed to compressed is approximately k: (N*k) / (N+k-2) ~ k(N/N) = k.

13. EVOLUTION AND SNOWFLAKES

Opponents of evolution often note that the chance of intelligent life is so small as to be impossible. Yet the same can be said of any given person (unique DNA or fingerprints) and any given snowflake. Once again the ambiguity in "the chance of this" appears.

14. CONCLUSION

What is unseen is often more important than what is seen. This is certainly true with coincidences. When we see an unlikely outcome that is memorable, we fail to see the many ways to generate the event in question and mistakenly conclude the coincidence cannot be chance since it seems so unlikely.

Be careful about statements involving chance grammar – they are often ambiguous. Is the event a pre-specified event (which is extremely unlikely) or is the event an after-the-fact event (which is generally much more likely)?

As Freeman Dyson noted, "The paradoxical feature of the laws of probability is that they make unlikely events happen unexpectedly often." (Oehlert, 2007).

Myers (2002) concluded, "That a particular specified event or coincidence will occur is very unlikely. That some astonishing unspecified events will occur is certain. That is why remarkable coincidences are noted in hindsight, not predicted with foresight."

A more precise summary would be, "That a particular pre-specified event or coincidence will occur at a pre-specified place and time is very unlikely. That some astonishing events noted after having occurred will occur somewhere sometime is certain. That is why remarkable coincidences are noted in hindsight, not predicted with foresight."

On the other hand, is it mere coincidence that in 1989 John Allen Paulos published *Innumeracy* and Diaconis and Mosteller published their seminal paper on coincidence in JASA? Is it mere coincidence that the last two digits of the publication year (89) match the number of letters in each of the author's names: eight for Fredrick and Diaconis, nine for John Allen? Is it mere coincidence that the sum of 8 and 9 just happens to equal the number of letters in the name of Fredrick Mosteller? Is it mere coincidence that $8 = 2^3$ while 9 is 3^2?

15. RECOMMENDATIONS

To be statistically literate, one must be very aware of the omnipresence of chance. The most common occurrence of chance in our everyday lives is through coincidences: unlikely events that are memorable.

We all need training to see what is unseen: to see the many possible ways that an observed outcome could have been produced. Only then can we appreciate the power of chance to explain coincidences.

More work is needed to demonstrate coincidences involving chance and determinism, and to decide whether a coincidence is really due to something other than chance.

REFERENCES

Blastland, Michael and Andrew Dilnot, (2009). *The Numbers Game: The Commonsense Guide to Understanding Numbers in the News, in Politics and in Life*. Gotham Books, Penguin Group. See http://www.StatLit.org/Blastland.htm.

Diaconsis, Persi and Fred Moesteller (1989). "Methods for Studying Coincidences." Journal of the American Statistical Association, 84:853-61. www-stat.stanford.edu/~cgates/PERSI/papers/ Mosteller89.pdf

Lesser, Larry (1999). Exploring the birthday problem with spreadsheets. *The Mathematics Teacher*. May 1999. Vol. 92, Iss. 5; pg. 407

Lord, Nick. (2010). "Binomial averages when the mean is an integer", The Mathematical Gazette 94, 331-332.

Marcus, Daniel A. (1998). Combinatorics: A Problem-Oriented Approach. The Mathematical Association of America.

Mathis, Frank (1991). A Generalized Birthday Problem. *SIAM Review* Vol 33, No2, pp. 265-270. www.jstor.org/pss/2031144

Myers, D (2002) Power of Coincidence. Skeptic Magazine. www.davidmyers.org/Brix?pageID=91

Neimark, Julie (2004). The Power of Coincidence. Psychology Today. www.psychologytoday.com/ articles/200407/the-power-coincidence

Oehlert (2007). Overreacting to patterns generated at random. www.statsmadeeasy.net/2007/02/

Paulos, John Allen (1989). *Innumeracy*.

Raymond, Robert and Milo Schield (2008). Numbers in the News: A Survey, 2008 ASA Proceedings of the Section on Statistical Education. [CD-ROM] P. 2848-2855. See www.StatLit.org/pdf/2008RaymondSchieldASA.pdf.

Schield, Milo (2005). Statistical Literacy and Chance. 2005 ASA Proceedings of the Section on Statistical Education. [CD-ROM] 2302-2310. See www.StatLit.org/pdf/2005SchieldASA.pdf.

Schield, Milo and Thomas V.V. Burnham (2008). Von Mises' Frequentist Approach to Probability, 2008 ASA Proceedings of the Section on Statistical Education. CDROM P. 2187-2194. See www.StatLit.org/pdf/2008SchieldBurnhamASA.pdf.

Schield, Milo (2008). Quantitative Literacy and School Mathematics: Percentages and Fractions, in *Calculation vs. Context: Quantitative Literacy And Its Implications for Teacher Education*. Edited by B. L. Madison and L. A. .Steen. 2008 Mathematical Association of America. p.87-107 At www.statlit.org/pdf/2008SchieldMAA.pdf or www.maa.org/Ql/cvc/cvc-087-107.pdf.

Schield, Milo (2010). Assessing Statistical Literacy: TAKE CARE in *Assessment Methods in Statistical Education: An International Perspective*. Edited by P. Bidgood, N. Hunt and F. Joliffe. Wiley Publishers, Ch. 11, p. 133-152.

Schield, Milo (2011b) *Statistical Literacy: Seeing Behind the Numbers*. 5th ed. Instant Publishers.

Schilling, Mark F (1990). The Longest Run of Heads. *The College Mathematics Journal*. Vol 21, Nr 3. http://gato-docs.its.txstate.edu/mathworks/DistributionOfLongestRun.pdf

Schilling, Mark F. (2012). The Surprising Predictability of Long Runs. Mathematics Magazine 85 (2012) 141–149.

Selkin, Lisa (2002). The Odds of That. The New York Times Magazine. http://nytimes.com/2002/08/11/magazine/11COINCIDENCE.html

ACKNOWLEDGEMENTS

Thanks to Blastland and Dilnot (2009) for provoking this last-minute paper with their idea of throwing rice to form random patterns on a floor. Thanks to Tom Burnham and Robert Raymond for their comments on earlier drafts. The author is at schield@augsburg.edu. These Excel spreadsheets are at www.statlit.org/Excel/2012Schield-Runs.xls, at www.statlit.org/Excel/2012Schield-Rice.xls and at www.statlit.org/Excel/2012Schield-Bday.xls and at www.statlit.org/Excel/2012Schield-Recursion.xls. Paper at www.statlit.org/pdf/2012Schield-MAA.pdf Slides www.statlit.org/pdf/2012Schield-MAA6up.pdf

Appendix A: Technical Details

First, are the runs just of one outcome (heads only) or do they include runs of both outcomes (heads and tails)? When the outcomes are binary, this issue is unique to coins. Normally, the event in question is so unlikely (chance of death due to lightning, a tornado, a hurricane, a tsunami, an earthquake, etc.) that the complement binary event cannot be viewed as a coincidence. But with coins, heads and tails are equally likely, so the chance of a run of four heads is just as likely as the chance of a run of four tails.

Whether to include both can be argued both ways.

- PRO: Coincidences involve unlikely events. The chance of getting four heads in four flips of a coin is just as likely (as unlikely) as getting four tails. They are both memorable coincidences and so both should be included.
- CON: With binary outcomes, normally one event is rare: much more unlikely than the other. The only run that is memorable involves the unlikely event. To include both outcomes in the case of coins may render the results inapplicable to other cases.

In this paper, runs of just one kind are counted since only the rarer of the two outcomes is typically involved in a coincidence.

Second, in N flips of a fair coin is a run of N-1 heads possible when a tail occurs between two of the heads? If the beginning and end terminate a run, then the answer is "no". Consider three flips of a fair coin. The only way to get a run of two heads is HHT or THH. But suppose we allow "wrap-around" so that the only way to terminate a run of heads is a tail. Now the sequence HTH can be said to involve a run of two heads.

Whether to allow "wrap-around" is an important issue when looking for the longest run. Consider the sequence: HHTHHH. Without wrap-around, the longest run is four heads. With wrap-around, the longest run is six heads. Allowing "wrap-around" increases the frequency of long runs.

Allowing "wrap-around" can be seen by arranging the coin outcomes in a circle instead of in a straight line. The only down-side of "wrap-around" occurs when all the outcomes are the same. In this case, the length of the run is unbounded.

Since our goal is to show that coincidences are much more frequent than expected, we show the results with wrap-around. But since students don't naturally think this way, we show the results without wrap-around as well.

Appendix B: Binomial Chance of No Successes

When the number of trials (N) equals 1/p where p is the probability of the desired outcome, then it can be shown that the probability of getting no desired outcome is always less than (1/e) but approaches (1/e) as N approaches infinity (P approaches zero). .

The chance of zero successes is given by $(1-p)^N$ where N = 1/p. So P(k=0) = $(1-p)^{(1/p)}$. Here is P(k=0) graphically using Excel.

Figure 16: P(k=0) vs. P when N=1/P

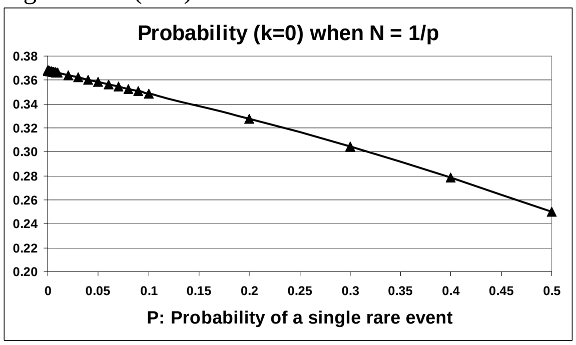
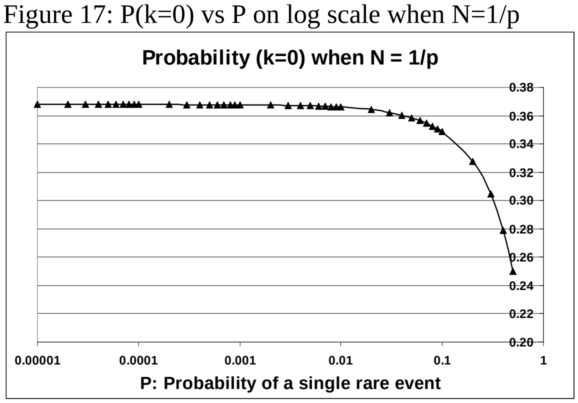


Figure 17: P(k=0) vs P on log scale when N=1/p



Note that as p approaches zero, P(k=0) approaches 0.36788. Note that 1/e equals 0.367888. It appears that as p approach zero, P(k=0) = 1/e.

Recall that Euler's number (e) is defined as $(1+1/N)^N$ as N approaches infinity. So, 1/e equals $(1+1/N)^{-N}$.

Dr. Robert Raymond (personal communication) pointed out that when the expected value (λ = Np) is fixed and N increases, the binomial distribution is approximated by the Poisson. When λ = 1 and k = 0, the Poisson is e^{-1}.

This convergence is noted in the Wikipedia article on Euler's constant: "e (mathematical constant)"

Appendix C: Binomial Chance of One Success

When N = 1/p, the chance of exactly one success is given by $(p/N)(1-p)^{N-1}$ *[nC1]. Since [nC1] = N, P(k=1) = $(1-p)^{(1/p-1)}$. = $(1-1/N)^{(N-1)}$. The following figures show P(k=0) and P(k=1) when N=1/p for various ranges of N. Notice the asymptotic approach in each figure.

Figure 18: P(k=0) and P(k=1): N < 100

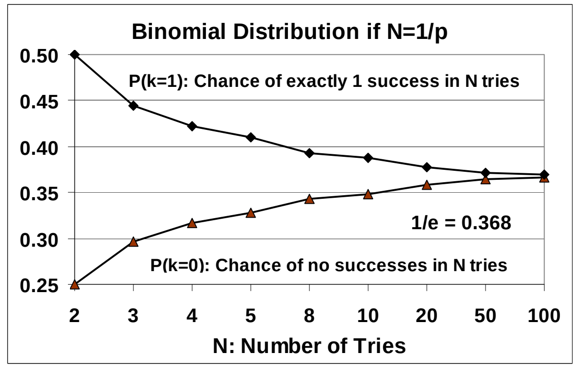
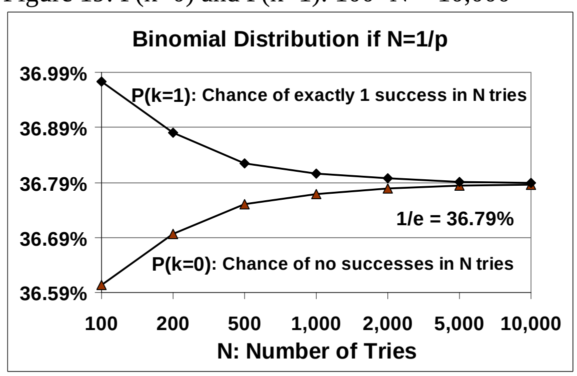


Figure 19: P(k=0) and P(k=1): 100<N < 10,000



The asymptotic approach to e^{-1} seems all but certain. The differences between P(k-0) = $(1-1/N)^N$, P(k=1) = $(1-1/N)^{(N-1)}$ and e^{-1} = Lim $(1+1/N)^{-N}$ seem to vanish as N approaches infinity.

This limit can be proven by using the Poisson approximation to the binomial. When λ = 1 and k = 1, the Poisson is e^{-1}.

Thus when Np = 1, P(k=0) and P(k=1) both approach e^{-1} as n increases. But P(k=0) approaches from below while P(k=1) approaches from above, so P(k=0) is never the modal value.

Appendix D: Distribution of Runs Given a Start

The phrase "distribution of runs" has three distinct meanings. (1) Distribution of run lengths given a specified starting position. (2) Distribution of the lengths of the longest run in N tries – in a string of length N. (3) Distribution of all runs in N tries – a string of length N -- where multiple runs are possible.

This section deals with coins using the first interpretation. Students readily agree when the starting position is fixed that longer runs are less likely than shorter runs; the longer the run, the less likely it is.

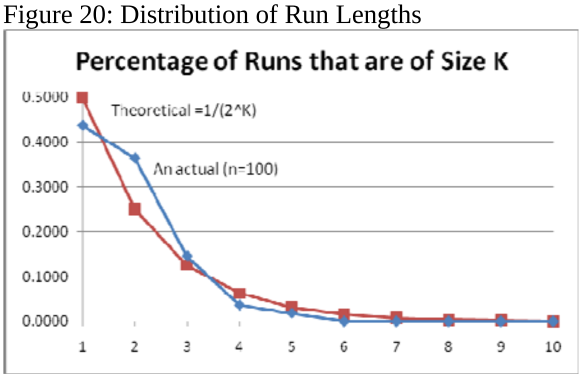
Given runs starting at a fixed position, consider the distribution of run lengths for a run of heads. This presumes there is a head in the first position and the random element is the distribution of heads in the following positions.

A run of length one has one chance in two since the first throw after the first head has a 50% chance of being a tail: P(T1 | H0) . A run of length two has one chance in four: P(H1, T2 | H0).

For a run of heads, the theoretical distribution for a run of length k starting in a given position is 1 chance in 2^k: P(Run length = k | H0) = 2^{-k}.

Here is an actual distribution of run lengths in flipping a fair coin 100 times:

Figure 20: Distribution of Run Lengths



It is tempting to say that the chance of a run of length one is 50%. But this abbreviated "chance-of" syntax is ambiguous as to the context. Review the three interpretations of that abbreviated phrase.

It is proper to say that the chance of a run of length one is 50% for a given starting point. A run of length zero is ignored as a contradiction in terms. .

The expected length of these runs is two as N approaches infinity. [It is 1.99 for N = 10.] The sum of k times 1/(2^k) is ½ plus 2/4 + 3/8 + 4/16 + 5/32 + 6/64, etc. This sum approaches two as shown in Table 7.

Table 7 Expected value of run length: k

k	2^k	P(k):1/2^k	K*(1/2^k)	Sum
1	2	0.5	0.5	0.5
2	4	0.25	0.5	1.0
3	8	0.125	0.375	1.38
4	16	0.0625	0.25	1.63
5	32	0.03125	0.15625	1.78
6	64	0.015625	0.09375	1.88
7	128	0.007813	0.05469	1.93
8	256	0.003906	0.03125	1.96
9	512	0.001953	0.01758	1.979
10	1,024	0.000977	0.009766	1.988
11	2,048	0.000488	0.005371	1.994
12	4,096	0.000244	0.002930	1.997
13	8,192	0.000122	0.001587	1.998
14	16,384	6.1E-05	0.000854	1.9990
15	32,768	3.05E-05	0.000458	1.9995
16	65,536	1.53E-05	0.000244	1.9997

Since run lengths of one are expected half the time, run lengths of two or longer are expected half the time and the median is 1.5. A run of length zero is ignored as meaningless. The mode is one.

This phrase, half the time, is easily misunderstood when the context is ambiguous. This does not mean that if we pick a particular cell, the chance that cell will mark the start of a run of length k is ½^k. In percent grammar this confuses:

1. 50% of the tries result in a run of length one.
2. 50% of the runs that start in a given position are of length one.

Of these two, the second is what is intended here. Only 25% of the tries at a given location result in runs of length one; 50% result in non-runs and 25% result in runs of length two or more.

Can this kind of exercise be done quickly in class with student-generated data and without any technology? One way is to ask each student to convert each of the last four digits of their ID number to *even* or *odd* and to represent *even* by zero and *odd* by one. Place a one (head) before each of these four digit numbers and look at the distribution of run lengths

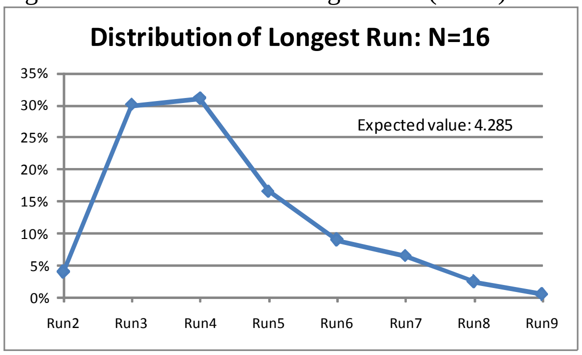
Note that this monotonically-decreasing distribution of runs by size is not the distribution of the longest run by size. The mean length for this distribution of runs stabilizes as the length of the series of flips increases. The mean length for the distribution of the longest run increases as the length of the series of flips increases.

This is where students get confused. They confuse the expected run length (two) with the expected length of the longest run.

Appendix E: Distribution of Longest Runs: #1

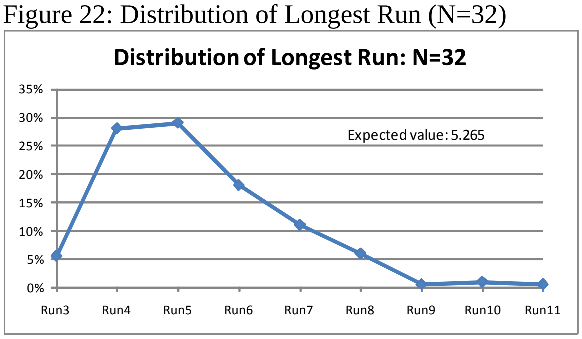
Consider the empirical distribution of the longest run length for various numbers of flips. In each case, the empirical data is obtained from 200 trials.

Figure 21: Distribution of Longest Run (N=16)



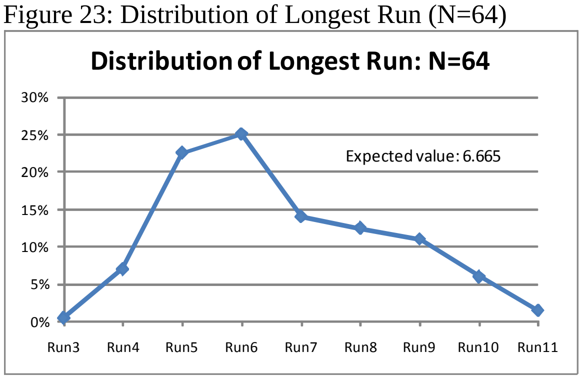
Note that $2^4 = 16$. Log (16) base 2 = 4. A run of four heads has a chance of one in 16 for runs starting at a given place. A run of four heads is expected when there are 16 tries – which are possible with N = 16 when wrapping the tail at the end around to the beginning.

Figure 22: Distribution of Longest Run (N=32)



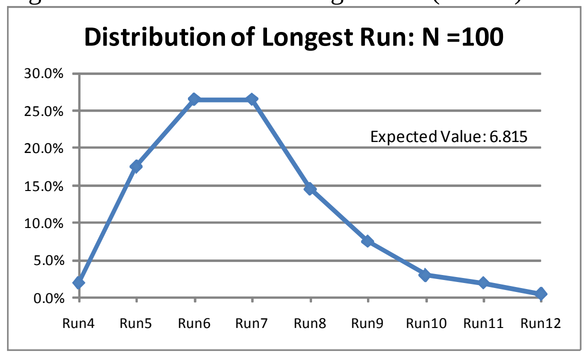
Note that $2^5 = 32$. Log (32) base 2 = 5. . A run of 5 heads has a chance of one in 32 starting at a given place. A run of 5 heads is expected when there are 32 tries – which are possible with N = 32 when wrapping the tail at the end around to the beginning.

Figure 23: Distribution of Longest Run (N=64)



Note that $2^6 = 64$. Log(64) base 2 = 6. A run of 6 heads has a chance of one in 64 starting at a given place. A run of 6 heads is expected when there are 64 tries – which are possible with N = 64 when wrapping the tail at the end around to the beginning.

Figure 24: Distribution of Longest Run (N=100)



Note that $2^7 = 128$. Log (127) base 2 = 7. A run of 7 heads has a chance of one in 128 starting at a given place. A run of 8 heads is expected when there are 128 tries – which are possible with N = 128 when wrapping the tail at the end around to the beginning.

Long runs are unexpected – a statistical coincidence. But students quickly realize that the greater the number of tries, the larger the longest expected run becomes.

The preceding discussion assumes that one wraps the data at the end around to the beginning so that N flips gives N different sequences for runs of any size up N.

Students may find it difficult to imagine this wrap-around. An alternative approach would be to add on an additional log(N) base 2 cells so that runs of the expected length have the full opportunity to be expressed.

Appendix F: Distribution of Longest Run: #2A

The theoretical distribution of the longest run in N flips of a fair coin for 2 ≤ N ≤12 was obtained using a "sledgehammer": generating all possible combinations and observing the longest run (k) for each unique combination.

Table 8 Longest Runs of Heads: 2 ≤ N ≤ 7

	N = Number of tries (length of string)					
k	2	3	4	5	6	7
0	1	1	1	1	1	1
1	2	4	7	12	20	33
2	1	2	5	11	23	47
3		1	2	5	12	27
4			1	2	5	12
5				1	2	5
6					1	2
7						1
Total	4	8	16	32	64	128
Mode	1	1	1	1	2	2
Median	1	1	1.5	2	2	2
Mean	1.00	1.38	1.69	1.94	2.16	2.34

Table 9 Longest Runs of Heads: 8 ≤ N ≤ 12

	N: Number of tries (length of string)				
k	8	9	10	11	12
0	1	1	1	1	1
1	54	88	143	232	376
2	94	185	360	694	1328
3	59	127	269	563	1167
4	28	63	139	303	653
5	12	28	64	143	315
6	5	12	28	64	144
7	2	5	12	28	64
8	1	2	5	12	28
9		1	2	5	12
10			1	2	5
11				1	2
12					1
Total	256	512	1024	2048	4096
Mode	2	2	2	2	2
Median	2	2	3	3	3
Mean	2.51	2.66	2.80	2.92	3.04

When flipping a coin, N = 6 is required to expect a longest run length of two (N = 5 for the median) while N = 12 is required to expect a longest run length of three (N=11 for the median). In this range, the expected value is close to k when $N = 2^k + (k-1)$.

See www.StatLit.org/Excel/2012Schield-Runs.xls

Appendix G: Distribution of Longest Run: #2B

Given the count distribution in Appendix F, one can readily obtain the probability of k being at least as big as the mean value.

Table 10 Probability k is at least the mean: 2 ≤ N ≤ 7

	N = Number of tries (length of string)					
k	2	3	4	5	6	7
0	0.25	0.13	0.06	0.03	0.02	0.01
1	0.5	0.5	0.44	0.38	0.31	0.26
2	0.25	0.25	0.31	0.34	0.36	0.37
3		0.13	0.13	0.16	0.19	0.21
4			0.06	0.06	0.08	0.09
5				0.03	0.03	0.04
6					0.02	0.02
7						0.01
Total	1.00	1.00	1.00	1.00	1.00	1.00
Mode	1	1	1	1	2	2
Median	1	1	1.5	2	2	2
Mean	1	1.38	1.69	1.94	2.16	2.34
P(k>=Exp)	0.75	0.38	0.5	0.59	0.31	0.37

Table 11 Probability k is at least mean: 8 ≤ N ≤ 12

N = Number of tries (length of string)					
k	8	9	10	11	12
0	0	0	0	0	0
1	0.21	0.17	0.14	0.11	0.09
2	0.37	0.36	0.35	0.34	0.32
3	0.23	0.25	0.26	0.27	0.28
4	0.11	0.12	0.14	0.15	0.16
5	0.05	0.05	0.06	0.07	0.08
6	0.02	0.02	0.03	0.03	0.04
7	0.01	0.01	0.01	0.01	0.02
8	0	0	0	0.01	0.01
Total	1.00	1.00	1.00	1.00	1.00
Mode	2	2	2	2	2
Median	2	2	3	3	3
Mean	2.51	2.66	2.80	2.92	3.04
P(k>=Exp)	0.42	0.46	0.51	0.55	0.30

Consider runs of length 2 so N = 5. Note that the chance that k is at least 2 is 59%. Now consider runs of length 3 so N = 10. The chance that the length of the longest run is at least three is 51%.

In both cases, when N = 1/P, the chance that the longest run is at least the expected value is greater than 50%: is more likely than not.

But any conjectures based on this limited range of N need to be tested for a much greater range of N.

Appendix H: Distribution of Longest Run: #3A

Schilling (1990) uses a neat recursive approach. Let $A_n(k)$ be the number of combinations with runs ≤ k where k is fixed and n varies. Consider k=3 in column E. For n = 0, 1, 2 and 3, $A_n(3) = (1/p)^n$. “For n > 3, each favorable sequence begins with either T, HT, HHT, or HHHT and is followed by a string having no more than three consecutive heads. Thus for n > 3,

$A_n(k=3) = A_{n-1}(3) + A_{n-2}(3) + A_{n-3}(3) + A_{n-4}(3)$”

To be unique, all $A_{n-1}(k)$ sequences can only be preceded by T; all $A_{n-2}(k)$ can only be preceded by HT, etc. for the k+1 terms.

Seeing why this is true is not obvious. To be true this procedure must generate sequences that are exclusive and exhaustive. Consider the specific sequences for k = 3 and n = 4.

$A_3(3)$: TTT, TTH, THT, THH, HHH, HTH, HHT and HHH. $A_2(3)$: TT, TH, HT, HH. $A_1(3)$: T, H.

Putting T before the $A_3(3)$ sequences gives TTTT, TTTH, TTHT, TTHH, THHH, THTH, THHT and THHH. Putting HT before the $A_2(3)$ sequences gives HTTT, HTTH, HTHT and HTHH. Putting HHT before the $A_1(3)$ sequences gives HHTT and HHTH. Putting HHHT before the $A_0(3)$ ‘sequence’ gives HHHT.

Note that all these four-character patterns are unique. Duplicates are impossible in this system. These four-character sequences are exhaustive: they include all the 16 possibilities except HHHH.

In this case, this procedure generated 15 four-character sequences that are exclusive and exhaustive.

Table 12 Cumulative Vertical Series: All runs ≤ k

	A	B	C	D	E	F	G	H
1		Length of the Longest run of heads in N tries						
2	N	0	1	2	3	4	5	6
3	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**
4	1	1	**2**	**2**	**2**	**2**	**2**	**2**
5	2	1	3	**4**	**4**	**4**	**4**	**4**
6	3	1	5	7	**8**	**8**	**8**	**8**
7	4	1	8	13	15	**16**	**16**	**16**
8	5	1	13	24	29	31	**32**	**32**
9	6	1	21	44	56	61	63	**64**
10	7	1	34	81	108	120	125	127

In a given column, the first number below the bold (in the n = k+1 row) is the sum of all the bolded numbers above. The formula is copied downward. Thus,

$A_1(0)$ B4: Sum(B3:B3), $A_2(0)$ B5: Sum(B4:B4), etc.
$A_2(1)$ C5: Sum(C3:C4), $A_3(1)$ C6: Sum(C4:C5), etc.
$A_3(2)$ D6: Sum(D3:D5), $A_5(2)$ D7: Sum(D4:D6), etc.
$A_4(3)$ E7: Sum(E3:E6), $A_4(3)$ E8: Sum(E4: E7), etc.

To get the non-cumulative distribution taken horizontally for k ≤ N, subtract adjacent cumulative counts within a given row Table 12 to get Table 13.

Table 13 Horizontal differences for a given n

A	B	C	D	E	F	G	H	I	J	K
k = Length of longest run of Heads in N tries										
n	Sum	0	1	2	3	4	5	6	7	8
0	**1**	1	0	0	0	0	0	0	0	0
1	**2**	1	1	0	0	0	0	0	0	0
2	**4**	1	2	1	0	0	0	0	0	0
3	**8**	1	4	2	1	0	0	0	0	0
4	**16**	1	7	5	2	1	0	0	0	0
5	**32**	1	12	11	5	**2**	1	0	0	0
6	**64**	1	20	23	12	**5**	2	1	0	0
7	**128**	1	33	47	27	**12**	5	2	1	0
8	**256**	1	54	94	59	28	12	5	2	1

The total row count, $(1/p)^n$, and the count in each cell are sufficient to generate row probabilities. See www.StatLit.org/Excel/2012Schield-Recursion.xls.

Note: Excel cannot handle these formulas for n greater than 1,023.

To see how this differs from simple Combinatorics, consider the row for n = 3. Here are the 8 combinations: k=0 (1): TTT. k=1 (4): HTT, THT, TTH, HTH. k=2 (2): HHT, THH. k=3 (1): HHH. Everything seems OK.

Now consider the row for n = 4. Here are the 16 combinations: k=0 (1): TTTT. k=1 (7): HTTT, THTT, TTHT, TTTH, HTHT, HTTH, THTH. k=2 (5): HHTT, THHT, TTHH, HHTH, HTHH. k=3 (2): HHHT, THHH. k=4 (1): HHHH

At this point it should be obvious that the formulas for combinations and permutations do not address this kind of problem. The distribution of the longest runs is a very different matter.

Appendix I: Distribution of Longest Run: #3B

Based on the Schilling recursion approach described in Appendix H, the following results are obtained. Schilling's approach agrees with the "sledgehammer" approach used in the previous appendices. More importantly, it allows a simpler extension for larger N. Table 14 gives summary statistics as a function of N.

Table 14 Longest Run of Heads: 1 ≤ N ≤ 1,023

N	Mode	Median	Mean
2	1	1	1.00
4	1	1	1.69
5	1	2	1.94
6	2	2	2.16
8	2	2	2.51
9	2	2	2.66
10	2	3	2.80
11	2	3	2.92
12	2	3	3.04
16	3	3	3.43
24	3	4	3.98
25	3	4	4.04
32	4	**4**	4.38
44	4	**5**	4.82
49	4	5	4.98
50	4	5	5.00
64	5	**5**	5.35
89	5	**6**	5.82
101	5	6	6.00
128	6	**6**	6.34
203	6	**7**	6.99
204	6	7	7.00
256	7	**7**	7.32
356	7	**8**	7.79
414	7	8	8.003
512	8	**8**	8.30
711	8	**9**	8.76
844	8	9	8.999
845	8	9	9.000
989	9	9	9.22
1,023	9	9	9.26

In each case, the median advances before the mean which advances before the mode as n increases.

Fitting a linear regression to the mean of the longest run as a function of the log of N base two gives Mean = -0.40 + 0.9647*Log_2(N), R^2 of 99.981%.

Fitting a linear regression to the median of the longest run as a function of the log of N base two gives Median= -0.5548 + 0.9846*Log_2(N): R^2 of 99.38%.

These values are close to Log_2(N) which supports using Log_2(N) as a rule of thumb on the upper limit of the mean or median length of the longest run.

Appendix J: Distribution of Longest Runs: 3C

Based on the Schilling recursion approach described in Appendix H, the distribution of the longest runs of heads is shown for selected values of N in the following tables. Table 15 shows the summary statistics for those values of N that correspond to N = $(1/2)^k$. [N=1,023 is shown in place of 1,024]

Table 15 Longest Run of Heads: 1 ≤ N ≤ 1,023

K	N	Mode	Median	Mean
3	8	2	2	2.51
4	16	3	3	3.43
5	32	4	4	4.38
6	64	5	5	5.35
7	128	6	6	6.34
8	256	7	7	7.32
9	512	8	8	8.30
10	1,023	9	9	9.26

In each case, the expected value (the mean) is less than k by up to 20%. But, this percentage difference decreases as N increases. For k = 10, the difference is less than 8%.

Table 16 shows the summary statistics for those values of N where the mean value of the distribution equals or just exceeds the integer values of k.

Table 16 Longest Run of Heads: 1 ≤ N ≤ 1,023

N	Mode	Median	Mean	Log_2(N)
12	2	3	3.04	3.58
25	3	4	4.04	4.64
50	4	5	5.00	5.64
101	5	6	6.00	6.66
204	6	7	7.00	7.67
414	7	8	8.00	8.69
845	8	9	9.00	9.72

The value of N required to give integer values for the mean lies between (1/p)^k and (1/p)^(k+1) and is less than (1/p)^(k+0.75) in this range.

For this case and range, a run of k is expected when N = (1/p)^(k+1) where "expected" identifies both the mean and the median. It is tempting to conjecture that this is generally true for other values of p and larger values of N. But without an analytic form for N as a function of p and k, this conjecture is simply a conjecture.

Schilling (2012) showed that for nq >> 1, the expected length of the longest run was given by

Exp(k) = Log(nq) base (1/p)

For coins, this reduces to log(n/2) base 2. This appears to be a lower limit for all values of n.

Appendix K: Combinations of k Tiles - Theory

Three theoretical approaches are presented here. One gives a lower limit for just straight-line clusters, a second gives a lower limit using poly-ominoes; the third gives an upper limit.

Consider just straight-line clusters starting with a given tile. There are eight ways to add a second tile to a first. There are two ways to add a tile to any of the existing patterns while maintaining a straight line. Etc. Thus the number of combinations relative to a given starting point is given by $8*2^{(k-2)}$.

A second approach drops the straight-lines requirement but does not allow touching on the points. The number of combinations of k tiles in two dimensions excluding diagonal touching has been estimated by $c*(lambda^k)/k$ where k is the number of tiles in the cluster, c is 0.3169 and lambda (L) is 4.0626. See the entry for "Polyomino" in Wikipedia.

A third approach uses branching. There are eight ways a second tile can be touching a first to form eight limbs. And there are at most 7 places that a third tile can be attached to the end of any of the eight limbs to form 56 branches. And there are at most 7 places that a fourth tile can be attached to the end of any of the 56 branches to give 392 leaves. The number of combinations given by this approach is $8*7^{(k-2)}$. This is definitely an upper-limit. It ignores duplicates and overlaps. For example the pattern N-E-S-W is the same as E-N-W-S. Limiting the number of leaves reduces the overlaps.

Results for all three methods are shown in Table 17.

Table 17 Limits: Number of Combinations of k Tiles

	Method 1	Method 2	Method 3
k	$8*[2^{(k-2)}]$	$c*(L^k)/k$	$8*[7^{(k-2)}]$
2	8	3	8
3	16	7	56
4	32	22	392
5	64	70	2,744
6	128	237	19,208
7	256	827	134,456
8	512	2,939	941,192
9	1024	10,615	6,588,344
10	2,048	38,812	46,118,408
11	4,096	143,342	322,828,856
12	8,192	533,814	2,259,801,992

Note that Methods 1 and 2 are lower limits while Method 3 is an upper limit.

Appendix L: Combinations of k Tiles - Empirical

An empirical approach counts how many unique shapes involving "runs" of size k can be places inside a square grid with sides L where $N = L^2$. A "run" involves any group of tiles that "touch" in any way.

Consider runs of size two in a 2x2 grid. There are six: two horizontal, two vertical and two diagonal.

Consider runs of size two in a 3x3 grid. There are 20 unique runs: 6 horizontal, 6 vertical and 8 diagonal. But runs of size two are uninformative; they are necessarily straight lines.

Consider runs of size three in a 2x2 grid. There are four: a single "L" shape rotated through four positions. Note that "L" is the only non-linear shape possible in a 2x2 grid. This situation is uninformative because it precludes straight-line runs.

Consider runs of size three in a 3x3 grid. There are 44 unique runs. There are eight straight-line runs: three horizontal, three vertical and two diagonal. There are 24 "L" shaped runs: four within each of the six overlapping 2x2 grids. And there are 24 new non-L shapes in a 3x3 grid. These shapes all involve a 2x3 part of the 3x3 grid. Suppose the nine cells are numbered starting at upper-left. Consider shapes 1-4-8 and 2-4-7: call them "5s" since their shape is close to that of clock hands at 5:00. . They are mirror-images when flipped, but not when rotated. Also consider 2-4-8: a "Y" shape. There are four ways each of these three shapes can be rotated for a total of 12 shapes. This situation is informative. It allows straight-line runs plus four other non-linear shapes. Note that this total of 44 unique runs is less than the 56 combinations estimated using Method 3.

Consider runs of size three in a 4x4 grid. There are 100 unique runs of size three in the four overlapping 3x3 grids and the nine overlapping 2x2 grids. There are 16 straight-line runs: four horizontal, four vertical and eight diagonal (two in each of the four 3x3 overlapping grids). There are 36 "L" shaped runs: four within each of the nine overlapping 2x2 grids. There are 48 non-L shapes: 12 (3*4) within each of the four of the 3x3 cell shaped runs. How can this exceed the 56 upper-limit in Table 17?

Going to runs of size four introduces new shapes and the problem quickly becomes increasingly complex. New mathematics is needed to describe the number of combinations possible in titles that can touch on their corners as well as one their sides. This study might be named "octi-ominoes" indicating eight ways to form a connection as opposed to the traditional "quadro-ominoes" which is limited to four sides.

Appendix M: B-day Problem: Chance of a Match

As mentioned previously, these results are approximate: they assume the trials are independent. Once the number of pairs or tries (T) is known, the chance of no match is $(1-p)^T$. The chance of at least one match is simply the complement: $1 - (1-p)^T$.

Table 18 shows the chance of a match for some common values of N – in multiples of five. Table 19 shows the number of people needed for selected percentages. Table 19 shows the approximate (memorable) chance of a match as a function of the number of people – in multiples of seven

Table 18 Birthday Problem: Chance of a Match #1

People	Pairs	Chance	Chance
N	**N(N-1)/2**	**No match**	**Match**
15	105	75%	25%
20	190	59%	41%
25	300	44%	56%
27	351	38%	62%
28	378	35%	65%
30	435	30%	70%
35	595	20%	80%
40	780	12%	88%
45	990	7%	93%
50	1,225	3%	97%

Table 19 Birthday Problem: Chance of a Match #2

People	Pairs	Chance	Chance
N	N(N-1)/2	No match	Match
15	105	75.0%	25.0%
23	253	50.0%	50.0%
26	325	41.0%	59.0%
27	351	38.2%	61.8%
28	378	35.5%	64.5%
30	435	30.3%	69.7%
32	496	25.6%	74.4%
41	820	10.5%	89.5%
48	1,128	4.5%	95.5%
59	1,711	0.9%	99.1%

Table 20 Birthday Problem: Chance of a Match #3

People	Pairs	Chance	Chance
N	N(N-1)/2	Match-1	Match-2
14	91	22.1%	20%
21	210	43.8%	40%
28	378	64.5%	60%
35	595	80.5%	80%
42	861	90.6%	90%
49	1,176	96.0%	95%
56	1,540	98.5%	99%
63	1,953	99.5%	99.5%
70	2,415	99.9%	99.9%

Appendix N: Three Meanings of Expected

"Expected" has three different meanings.

1. *Expected* is what is most likely (the mode). This is an ordinary or colloquial meaning.
2. *Expected* is what happens most (a majority) of the time: it is more likely than not. In ordinary usage, little attention is paid to the difference between most of the time (majority) and most likely (mode). In any process with just two outcomes, the most likely is what happens most of the time (the majority). There must be at least three outcomes in order for there to be a difference between majority and mode.
3. *Expected* is what happens on average (the mean). This is the technical or statistical meaning. : The "expected" value is the average value of a probability distribution.

   Consider the binomial distribution of a Bernoulli random variable with p chance of success in a single try. The expected number of successes in N tries is given by N*p.

   If this expected value is an integer, the mean, median and mode coincide. (Lord, 2010)

   Although the mode is most likely, this does not mean the mode is more likely than not (occurs most of the time). As shown in Appendix C, the probability of k=1, P(k=1), never exceeds 50%. So the mode does not occur most of the time.

In ordinary usage, an event is *expected* in N tries if it has one chance in N. Which of these three usages does this usage entail?

Consider a six-sided die where one of the six sides is labeled a success. Rolling one success is expected when rolling six of these die. This situation satisfies the criteria a Bernoulli random variable. When N = 1/p, the associated binomial distribution has a mean, median and mode of one. So, one is both the long-term average (mean) and the most likely outcome (mode) – but it seems it is not what happens most of the time (majority).

Since the median always equals the mean when N = 1/p, having the expected number of success *or more* always happens most of the time. Using this extended interpretation of *expected,* one can say the expected (or more) happens most of the time.

If *expected* is extended to include the average or higher for most of the time, then when N = 1/p, all three criteria are satisfied: average (mean), most likely (mode) and most of the time (majority).