Humanistic Mathematics Network Journal

Issue 21 Article 20

12-1-1999

Tea Tasting and Pascal Triangles

Paul Alper
University of St. Thomas

Yongzhi Yang
University of St. Thomas

Tea Tasting and Pascal Triangles

Paul Alper
QMCS Department
University of St. Thomas
St. Paul, MN 55105
e-mail: p9alper@stthomas.edu

Yongzhi Yang
Mathematics Department
University of St. Thomas
St. Paul, MN 55105
y9yang@stthomas.edu

INTRODUCTION TO THE PHYSICAL SITUATION

R. A. Fisher is generally considered to be the most famous statistician of all time. He is responsible for such familiar statistical ideas as analysis of variance, Fisher's exact test, maximum likelihood, design of experiments and hypothesis testing, among others. He was very prolific, but as James Newman kindly put it, "Fisher is not an easy writer." [1] Even today, almost four decades after his death, people on opposing sides of a statistical argument can quote Fisher to prove their respective points. Fisher was indeed very interested in what might be termed the philosophy of statistics; that is, to paraphrase the famous aphorism in management science, how to do things right statistically and how to do the right statistical things.

Fisher, when asked what he did for a living, would reply he was a scientist because he felt that statistics more properly belonged to the sciences and not to mathematics. Indeed, he was never a professor of statistics but instead was a professor of eugenics (at London University)–the overwhelmingly negative connotations of eugenics came years later with the rise of nazism–and then a professor of genetics (at Cambridge University). Despite his reputation as someone difficult to interpret unambiguously, one of his philosophical contributions to the annals and history of the discipline is both well known, and, according to Newman, "a model of lucidity and required no mathematics other than elementary arithmetic. It demands of the reader the ability to follow a closely reasoned argument, but it will repay the effort by giving a vivid understanding of the richness, complexity and subtlety of modern experimental method."

Fisher, in his *Design of Experiments*, entitled this contribution "The Principles of Experimentation, Illustrated by a Psycho-Physical Experiment." [2] Newman, in his four-volume *magnum opus, The World of Mathematics*, reprinted it in its entirety under the name by which it is perhaps more commonly known: "Mathematics of a Lady Tasting Tea." Fisher uses the unlikely example of a woman who claims to be able "to discriminate whether the milk or the tea infusion was first added to the cup" in order to elucidate which design principles are essential and which others are but auxiliary. He discusses why she should be informed of how many cups of each there are and why three cups of milk first and three cups of tea first would not be convincing even should she be able to get all six correct: with three and three, the chance of identifying all six merely by guessing would be 1 in 20 which is *not less than .05*–to Fisher, "It is usual and convenient for experimenters to take 5% as a standard level of significance." Whereas, with four of each, the chance of identifying all eight merely by guessing would be 1 in 70 which *is* less than .05. Fisher could have set this up within his context of significance testing where the null hypothesis would be

H_0: The number of correct answers is due to chance alone.

He further goes on to show that five cups of one and three of another would be inferior in the sense that guessing all correct in this instance would be 1 in only 56 and thus not as impressive. Fisher also points out that letting "the treatment of each cup be completely determined by chance, as by the toss of a coin, so that each treatment has an equal chance of being chosen" while increasing the "sensitiveness" of the experiment to 1 in 256, should be foregone because it might bewilder the subject and "deny her the real advantage of judging by comparison." In this last instance, the null hypothesis would be

H_0: Probability of success for any one cup is .5

Fisher also looked at other refinements such as avoiding or limiting possible confounding elements so as to ensure that the tea, milk and temperature were the same for each cup; however, perhaps his greatest emphasis in this imaginary psycho-physical experiment is on the need for randomization, a concept, taken for granted today by all statisticians, which back in the 1930s when this was written, necessitated proselytizing.

The usual way this problem in all its various forms is analyzed is via combinatorics, a subject which is often difficult for beginners. Combinatorics frequently result in formulas which are sufficiently obtuse so that insight is hard to come by. Fortunately, this problem can be attacked via simulation. Furthermore, Fisher's tea tasting experiment in all its various forms leads to Pascal triangles, a subject more conducive to insight and one which has a long history in mathematics and seems to pop up in unlikely places. [3] Because the resulting Pascal triangle is most familiar for the case in which the treatment of each cup is determined by *chance*, that will be the first one simulated. The next simulation will be for *equal* cups of milk first and tea first; it will be seen that a "quadratic" Pascal triangle results. The last simulation will be for the *unequal* (but known cups of each) case; this too will result in a type of Pascal triangle.

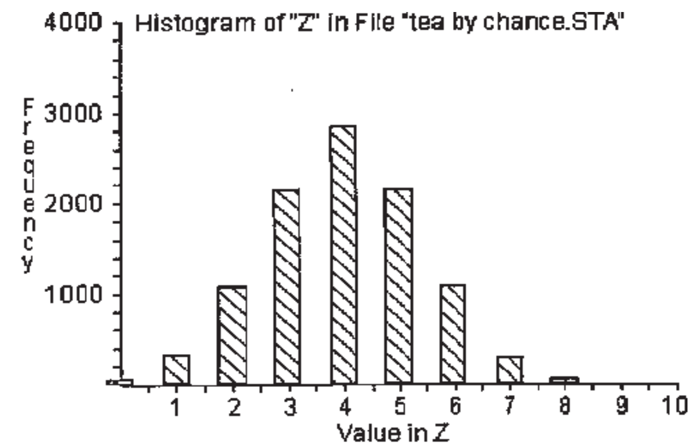
SIMULATION AND PASCAL TRIANGLES

With the ubiquity of computers, one is tempted to avoid messy combinatoric formulas by simulating the experiment, and thus not only obtaining a numerical answer, but, in addition, acquiring some clue about the underlying structure of this problem. The simulations will be carried out in *Resampling Stats*, a software package designed to do simulation in statistics. [4] The simulation could be carried out in any computer language, but what is taking place is a computerization of what one could do physically; computers, of course, merely speed things up.

NUMBER OF CUPS OF TEA OR MILK DETERMINED BY CHANCE

When the treatment of each cup is determined by chance, the following simple program could be used when the lady is presented with eight cups and is not informed of how many are milk or tea first. The "sample" command means sample with replacement so that while "deck" contains eight elements, four 1's followed by four 2's, "deck$" has eight elements with from zero to eight 1's depending on chance. The "score" command is used to keep track of her number of successes in the 10,000 trials; at the end of the "repeat" loop, the variable "z" has 10,000 elements.

```
A lady testing tea "by chance" example
concat 4 #1 4 #2 deck
maxsize z 15000
repeat 10000
  sample 8 deck deck$
  count deck$ = 1 a
  score a z
end
histogram z
count z = 8 k
divide k 10000 kk
print kk
```



KK = 0.0044

The symmetry in the histogram is clearly evident, and any general result ought to exhibit this property. Furthermore, the program is easily modified to deal with any number of cups merely by changing the "sample" command; of course, by changing the "count" command's equal sign to greater than or equal for the variable "z," one could find the probability of more than a certain number of successes instead of exactly that many. In contrast, the general combinatoric formula for the probability of getting l or more correct out of the n chances when the probability of success is .5 is

$$\sum_{k=1}^{n}\binom{n}{k}(0.5)^{n}$$

This formula, all too often, fails to reach the beginning student. As it turns out, this (binomial) formula is related to Fibonacci numbers and can be displayed as a Pascal triangle which is often shown in an isosceles shape:

```
                1
               1 1
              1 2 1
             1 3 3 1
            1 4 6 4 1
          1 5 10 10 5 1
        1 6 15 20 15 6 1
      1 7 21 35 35 21 7 1
    1 8 28 56 70 56 28 8 1
```

Another, but less familiar way to show the Pascal triangle is to display it as a right triangle.

n/*k*	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1

Here, *n* stands for the number of cups which the lady tastes and *k* is the number of correct choices. For example, the probability of getting 6 or more correct out of 8 chances is given by using the proper row of either of the Pascal triangles

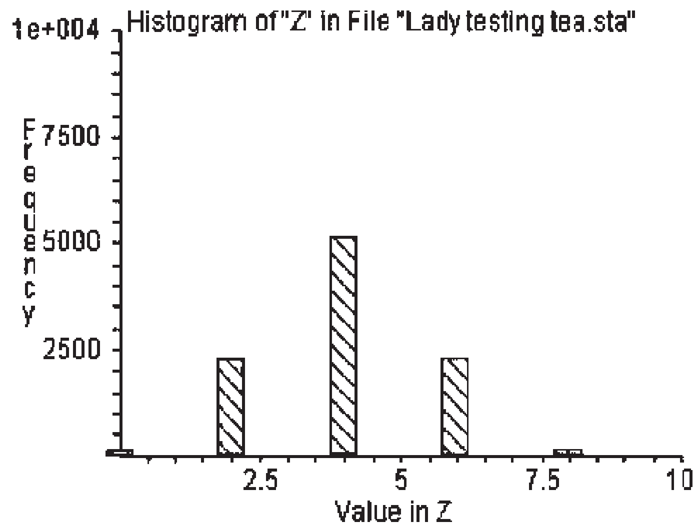
$$\frac{(28+8+1)}{(1+8+28+56+70+56+28+8+1)} =$$

$$\binom{8}{6}(0.5)^8 + \binom{8}{7}(0.5)^8 + \binom{8}{8}(0.5)^8 = \frac{37}{256}$$

FISHER'S ORIGINAL SITUATION

In Fisher's *original* proposal the lady is informed that there is an equal number of milk or tea first–four of each in his article. The simulation given below is very similar to the *by chance* program already given. The main difference is that the "sample" command is replaced by "shuffle" and "take" which means that sampling is done without replacement.

```
A lady testing tea example
concat 4 #1 4 #2 deck
maxsize z 15000
repeat 10000
  shuffle deck deck$
  take deck$ 1,4 milk
  take deck$ 5,8 tea
  count milk = 1 a
  count tea = 2 b
  add a b c
  score c z
end
histogram z
count z = 8 k
divide k 10000 kk
print kk
```



Once again the histogram indicates symmetry. Further, only an even number of correct choices is indicated as possible. The combinatoric formula for obtaining *a* correct cups of milk (and as it turns out, therefore *a* correct cups of tea) when there are *m* cups of milk first and *n*=*m* cups of tea first is:

$$\frac{\binom{m}{a}\binom{n}{a}}{\binom{2n}{n}}$$

So to speak, this combinatoric formula lacks transparency. That is, while it is correct, insight is hard to come by. However, coupled with the simulation results there is a suggestion of some deeper structure. That deeper structure turns out to be what might be called a "quadratic" Pascal triangle:

$$
\begin{array}{c}
1^2 \\
1^2 \quad 1^2 \\
1^2 \quad 2^2 \quad 1^2 \\
1^2 \quad 3^2 \quad 3^2 \quad 1^2 \\
1^2 \quad 4^2 \quad 6^2 \quad 4^2 \quad 1^2 \\
1^2 \quad 5^2 \quad 10^2 \quad 10^2 \quad 5^2 \quad 1^2
\end{array}
$$

That is, each element of the usual Pascal isosceles triangle is squared. Or, the ordinary Pascal isosceles triangle is "multiplied" by itself. The same applies to "multiplying" right angle Pascal triangles to obtain:

n/k	0	1	2	3	4	5
0	1^2					
1	1^2	1^2				
2	1^2	2^2	1^2			
3	1^2	3^2	3^2	1^2		
4	1^2	4^2	6^2	4^2	1^2	
5	1^2	5^2	10^2	10^2	5^2	1^2

where there are n cups of milk first and $n=m$ cups of tea first; $2k$ is the number of incorrect choices. Hence, the solution for any equal number of milk or tea first may be found by using the corresponding row of a (quadratic) Pascal triangle. For example, when there are four cups each, the probability of getting 6 or more correct, equivalent to obtaining 2 or fewer failures is analytically given by

$$\frac{(1^2+4^2)}{(1^2+4^2+6^2+4^2+1^2)} = \frac{17}{70}$$

When there are five cups each, we just move down one row. The Pascal triangle gives a neat pictorial presentation of how much more difficult it is for the lady to guess correctly; the combinatoric formula does not reveal this so readily.

UNEQUAL NUMBER OF KNOWN CUPS OF EACH

When m is less than or equal to n but still known to the lady, the program just given is easily modified; the only substantive difference is the "deck" and the "take" commands utilize the particular m and n. The combinatoric formula, on the other hand, is quite frightening to beginners:

$$\frac{\binom{m}{a}\binom{n}{n-m+a}}{\binom{m+n}{m}}$$

The combinatoric formula and the simulation suggest some structure in this situation as well. However, for this situation of unequal m and n, things aren't quite so tidy. Nonetheless, some sort of "multiplication" is taking place. For example, when there is one more tea first than milk first, $n=m+1$, the following triangle results:

$$
\begin{array}{c}
1 \\
1 \quad 2 \\
1 \quad 6 \quad 3 \\
1 \quad 12 \quad 18 \quad 4 \\
1 \quad 20 \quad 60 \quad 40 \quad 5
\end{array}
$$

Due to the lack of symmetry, this is better viewed via the right angle Pascal triangles.

m/k	0	1	2	3	4
0	1				
1	1	2			
2	1	6	3		
3	1	12	18	4	
4	1	20	60	40	5

(1)

where m is the number of cups of milk first and n is the number of cups of tea first. $2k$ is the number incorrect and $2m-2k+1$ is the number correct.

The element in the ith row in the resulting triangle in (1) is obtained from multiplying the element in the ith row and the element in the same column in the $(i+1)$th row in the ordinary right angle Pascal triangle. For example, when $m=2$ and $n=3$, the third row in (1) is constructed by multiplying the element in the third row and the element in the same column in the fourth row in the Pascal triangle.

$$
\begin{array}{cccc}
1 & 2 & 1 & \\
\times & \times & \times & \\
1 & 3 & 3 & 1 \\
\Downarrow & \Downarrow & \Downarrow & \\
1 & 6 & 3 &
\end{array}
$$

For example, when the difference between n and m is two, then the following triangle results:

m/k	0	1	2	3	4
0	1				
1	1	3			
2	1	8	6		
3	1	15	30	10	
4	1	24	90	80	15

(2)

where m is the number of cups of milk first and n is the number of cups of tea first. $2k$ is the number incorrect and $2m-2k+2$ is the number correct.

The element in the ith row in the resulting triangle in (2) is obtained from multiplying the element in the ith row and the element in the same column in the $(i+2)$th row in the Pascal triangle. In general, the triangle which results for any m and n is to multiply the ith row by the $(i+n-m)$th row.

The lack of symmetry when m is unequal to n implies that Pascal triangle approach loses some of its visual advantage over the combinatoric formula. Nevertheless, the Pascal triangle still displays some pictorial benefit.

PHILOSOPHICAL CONCLUSIONS

Being a scientist, Fisher's main purpose in this tea tasting scenario was to illustrate the ideas behind the design of experiments when psychology was combined with the physical. On the other hand, mathematicians often have a different agenda such as showing surprising and non-intuitive interconnections. That the tasting of tea as described by Fisher should lead to a quadratic Pascal triangle is esthetically pleasing to a mind with a mathematical bent. Just as important is the fact that when Fisher's scenario is altered to allow unequal (but known total cups of each), the Pascal triangle can be easily used to determine numerical results as the number of cups change; this is in contrast to the combinatoric formula which tends to hide what is taking place and is often difficult to calculate numerically.

BIBLIOGRAPHY

1. Fisher, R.A., "The Mathematics of a Lady Tasting Tea," *The World of Mathematics*, edited by Newman, J.R., pp 1512-1521, 1456-58, 1956.
2. Fisher, R.A., "The Principles of Experimentation, Illustrated by a Psycho-Physical Experiment," *Design of Experiments*, Edinburgh, Oliver and Boyd, pp 11-25, 1949.
3. Bondarenko, B.A., *Generalized Pascal Triangles and Pyramids; Their Fractals, Graphs, and Applications*, Fibonacci Association, 1993.
4. Simon, J.L., *Resampling: The New Statistics*, Resampling Stats, Inc, 1997.

Roots

Don Pfaff
Math Department, University of Nevada, Reno

I think that I shall never see
The square root of the number three,
A number so irrational
It cannot be conceived at all.
Squares were made for fools like me,
But only God can root a three.