Statistical paradoxes: when maths and logic collide with one another

Azzam F. G. Taktak (Department of Medical Physics and Clinical Engineering, Royal Liverpool University Hospital) has some examples of conditional probability

"By swapping to door 2, you actually double your chances of winning"

The world of statistics is full of paradoxes which have generated a lot of interest and debate for many years.[1] One of the main themes that these paradoxes take are related to the topic of probability and, more specifically, conditional probability. These paradoxes demonstrate quirks which have significant implications on data analysis in research. The literature is full of examples were researchers fell foul of the pitfalls that these paradoxes illustrate. Ben Goldacre's *Bad Science* book[2] is an excellent reference highlighting how such pitfalls can lead to wrong, or simply dubious, conclusions. I have picked out a few examples here which make excellent party material for sad physicists and engineers like myself (I am available for weddings, christenings and other occasions). I have linked these examples to some real-world medical applications.

THE TWO ENVELOPES PARADOX

I have two identical envelopes containing cash. The amount of cash in one envelope is twice the amount in the second. I pick one envelope and I have the chance to open it or swap for the second. I reason that if the amount in the envelope in my hand is X, the amount in the other envelope is either $2X$ or $X/2$, each having a probability of 1/2. The expected value of the amount in the second envelope is therefore:

$$\frac{1}{2}\times 2X + \frac{1}{2}\times\frac{X}{2} = \frac{5}{4}\times X$$

which is higher than X. It is therefore in my advantage to swap. But then I apply the same reasoning to the second envelope and swap again, and go on swapping forever. Where did I go wrong? The problem is in my assumption that the amount in the envelope in my hand is X and making my calculations based on that assumption. However, X is a random value so $1.25X$ is also a random value.

Another way to look at it is that we were told that one envelope contains twice the amount of the second. We therefore defined the sample space at the start of the problem as $\{X, 2X\}$. Once we made a prior assumption about the first envelope (which we also called X just to be confusing), we based our assumption for the second envelope relative to the first and our sample space is now $\{X/2, 2X\}$. This sample space violates the condition of the original question as $2X$ is four times $X/2$ and not twice. The correct way to look at it is that each envelope has either X or $X/2$ with a probability of 1/2 so the expected value of each envelope is:

$$\frac{1}{2}\times X + \frac{1}{2}\times 2X = \frac{3}{2}\times X$$

This is certainly true. If, for example, one envelope contains £10 and the other contains £20 ($X = 10$) and if I do this experiment 1,000 times, then on average I will gain £15 per experiment.

THE MONTY HALL PARADOX

This is an interesting paradox on conditional probability. Supposing you were in a game show and the host shows you three doors; 1, 2 and 3. Behind one of these doors is a car. You pick a door at random, say 1. Before the host opens door 1, he says 'let me open another door and show you that it does not contain the car', just to add suspense. He then opens door 3 and shows that it does not contain the car. He then gives you the option to stay with door 1 or swap to door 2. What should you do? The most logical answer is that the probability of the car being behind door 1 is equal to the probability of it being behind door 2, which is 1/2. Wrong! By swapping to door 2, you actually double your chances of winning.

Here is why. Each door has a probability of 1/3 of containing the car. Supposing that the car was indeed behind door 1. The host can open either door 2 or door 3, each with a probability of 1/2. The total probability of the car being behind door 1 and the host opening door 2 or door 3 is:

$$1/3 \times 1/2 + 1/3 \times 1/2 = 1/3$$

In this case, if you swap you will lose. Now, if the car was behind door 2, the host can only open door 3 because if he opens door 2 he will reveal the car. The total probability of the car being behind door 2 and the host opening door 3 is therefore $1 \times 1/3$. In this case, if you swap you will win. The same argument holds if the car was behind door 3. There is therefore a 2/3 chance that you will win if you swap and a 1/3 chance that you will lose.

To put it mathematically, if the probability of the car being behind door 2 is $P(C = 2)$, the probability of you selecting door 1 is $P(S = 1)$, the probability of the host opening door 3 is $P(H = 3)$, using Bayes theorem:[3]

$$P(C=2 \mid H=3, S=1) = \frac{P(H=3 \mid C=2, S=1)\times P(C=2 \mid S=1)}{\sum_{n=1}^{3} P(H=3 \mid C=n, S=1)\times P(C=n \mid S=1)}$$

$$= \frac{1\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3} + 1\times\frac{1}{3} + 0\times\frac{1}{3}} = \frac{2}{3}$$

Here again, the original sample space is {1/3, 2/3}, the probability of winning or losing. The host did not open a door at random. He opened a door knowing that it is not the one you picked and that it does not contain the prize, so the original sample space remains unchanged. ▶

▶ **FIGURE 1.** A simulation program for the boy or girl paradox.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   Boy or Girl Paradox Simulation Program             %
%   Author: A. Taktak                                  %
%   Date: 4 July 2010 (this is how I spend my weekends ☹) %
%   Use: At the command line, type:                    %
%          > p=boyorgirl(guess)                        %
%       where guess is either 0 (for boys) or 1 (for girls) %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function output=boyorgirl(guess)
correct=0;

for n = 1:1000
  % Sample a random number from binomial distribution with a probability
  % of 0.5 for each child
  ChildA=binornd(1,0.5,1,1);
  ChildB=binornd(1,0.5,1,1);

  % Ensure that at least one of the children is a boy as the stated in the problem
  while ((ChildA~=0)&&(ChildB~=0))
    ChildA=binornd(1,0.5,1,1);
    ChildB=binornd(1,0.5,1,1);
  end

  % The next condition is important since we don't know which of the two
  % children is a boy. Note that if the first condition is true for
  % Child A then we skip the condition for Child B so we are not cheating
  % by doubling the chances of a correct answer
  if ChildA==0
    if ChildB==guess
      correct=correct+1;
    end
  else
    if ChildB==guess
      correct=correct+1;
    end
  end
end

% express the output as a probability
output=correct/1000;
```

▶ **FIGURE 2.** Calculating the relative risk of having a car accident whilst drunk using hypothetical data (http://clinengnhs.liv.ac.uk/MedStats/MedStats_Demos.htm).

Relative Risk and Odds Ratio

Exposure	Outcome	
	Yes	No
Yes	300	50
No	700	950

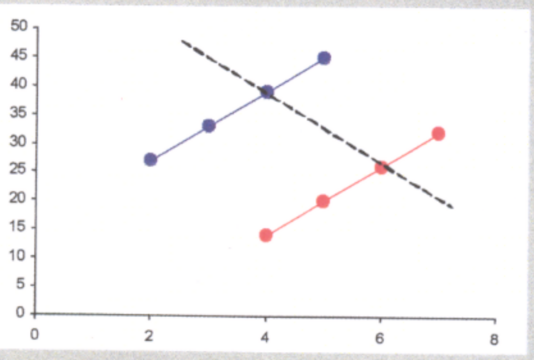
Calculate

Relative Risk (RR) (95% CI)= 2.02 (1.88 - 2.17)

Odds Ratio (OR) (95% CI)= 8.14 (5.94 - 11.16)

Chi-squared test for no association: 216.45 (p: 0.00)

p > 0.1	little evidence of an association
0.1 ≥ p > 0.05	weak evidence of an association
0.05 ≥ p > 0.01	moderate evidence of an association
p ≤ 0.01	strong evidence of an association



▶ **FIGURE 3.** Simpson's paradox illustration.

BOY OR GIRL PARADOX

This is another interesting one. Consider a family that has two children. If one of the children is a boy, what is the probability that the other one is also a boy? Here again the most logical answer is 1/2. Wrong again! The correct answer is 2/3.

The reason is that there are three possible combinations for a family with two children. The sample space is: {BB, GG, BG}.

We were told that one of the children is a boy so it cannot be GG. The sample space therefore reduces to: {BB, BG}.

We were told that one of the children is a boy so we are drawing one of the children in a pair at random; not the pair. There are three boys in the above two combinations, each having a probability of 1/3 of being picked at random. For two of them, the probability of having a brother is 1 and for one it is 0, hence the total probability:

$$1\times\frac{1}{3}+1\times\frac{1}{3}+0\times\frac{1}{3}=\frac{2}{3}$$

Note that this is not the same as the problem where 'if a family has one child who is a boy, what is the probability of the next child also being a boy?'. In this case, it *is* 1/2.

If you do not believe me, try the simulation program I wrote in MATLAB shown in figure 1. Boys are represented by 0 and girls as 1. If you enter your guess as 0, you will get an answer close to 0.67. If you enter 1, you will get close to 0.33.

THE LOVER'S DILEMMA

A man who lives in Midville has two lovers, one in Eastville and one in Westville (he is obviously not a clinical scientist or he would be far too busy for that sort of thing). There are an equal number of trains that go to Eastville than there are to Westville. In order to decide which lover he should visit, he arrives at the train station at random times of the day and at random days of the week and takes the next available train. He expects that after a long period of time doing this, he would see his two lovers an equal number of times (another reason to suspect that he is not a clinical scientist). However, he finds out that he is five times more likely to see the lover in Eastville than the one in Westville. Why is that?

The reason is quite simple. Although he arrives at the train station at random times, trains do not arrive at random times. They have a timetable to follow. Suppose that trains to Eastville arrive once an hour, on the hour. The trains to Westville also run once an hour but arrive at 10 minutes past the hour. There is therefore a 50-minute window every hour that the next train is the one heading to Eastville but only a 10-minute window for the train to Westville. The sample space is therefore {1/6, 5/6}.

This is known in statistical terms as the class prior. It has a huge implication on statistical modelling. Supposing I have a dataset from a particular clinic, say a diabetes clinic, and I am looking for a particular event, say heart failure. I use a statistical model (like logistic regression) to predict the probability of heart failure using a range of risk factors like age, BMI, family history, smoking, etc. If 50 per cent of the patients from that clinic went on to develop heart failure, i.e. a class prior of 0.5 because they are at high risk, my model has a 50 per cent overall chance of predicting heart failure. It would be wrong of me then to use the same model on a random person off the street since the class prior for the general public (also known as prevalence) is a lot less than 0.5.

A TOMMY COOPER JOKE

Last Christmas, I got a Tommy Cooper joke book from my wife which I thoroughly enjoyed. One joke of particular relevance to the topic here is as follows. It is said that three out of every ten car accidents are caused by drunk drivers. That means that seven out of ten are caused by sober drivers so if everyone gets drunk, there would be fewer car accidents. At the risk of killing the joke, I will try to show the flaw in this argument.

If we represent the event of being drunk as D, the event of being sober as S and the event of having a car accident as A, then:

$P(D \mid A) = 0.3$
$P(S \mid A) = 0.7$
i.e. $P(D \mid A) < P(S \mid A)$.

This is not the same as:

$P(A \mid D) < P(A \mid S)$

which is what the joke implies. This is known as the transposition of conditioning fallacy. The only way to test the inequality is by doing a controlled trial where we choose 100 drunk drivers and 100 sober drivers, send them out on a long journey and see how many of them make it home safely. Clearly not feasible. The other option is to use observational data (more about this in the next section). If we assume that the above statistics came from a sample of 1,000 car accident records, we send out anonymous questionnaires to 1,000 drivers at random who have not been involved in car accidents and ask them if they have ever been under the influence of alcohol whilst on the wheel. Suppose that 95 per cent of respondents said no (I think it is safe to assume that at least 95 per cent of drivers on the road are sober). That means the relative risk of having a car accident whilst drunk is about 2 (figure 2). That is, there is 100 per cent more risk of being involved in a car accident whilst drunk than if you were sober. Sorry Mr Cooper!

SIMPSON'S PARADOX

Simpson's paradox is a phenomenon that occurs when studying correlation between two variables from observational data without taking the effect of confounders into consideration. Edward H. Simpson first described this phenomenon in a technical paper in 1951.[4] This effect can be best illustrated in figure 3. Suppose variable Y has a positive linear relationship with variable X. The slope of the best fit line is constant but the intercept depends on the level of a binary variable C so that when $C = 0$, the correlation is described by the blue line and when it is 1, it is described by the red line. Now if we did not take C into account, we might do an experiment which results in the dashed black line which has a negative slope so we might wrongly conclude that X and Y have an inverse relationship.

In observational (non-randomised) studies comparing treatments, it is likely that the initial choice of treatment would have been influenced by patients' characteristics such as age or severity of condition, so any difference between treatments could be accounted for by these original factors. To take a real-world example, Charig *et al.* undertook a historical comparison of success rates in removing kidney stones.[5] He showed that open surgery had a success rate of 78 per cent (273/350) while keyhole surgery had a success rate of 83 per cent (289/350). The study concluded that keyhole surgery improves the chances of a successful outcome. However, when stone diameter was taken into account, this showed that, for stones < 2 cm diameter, 93 per cent (81/87) of cases with open surgery were successful compared with just 87 per cent (234/270) of cases with keyhole surgery. One would naturally assume that for larger stones, keyhole surgery must perform much better than open surgery in order to make up the difference. In fact it was observed that for stones of > 2 cm, success rates were 73 per cent (192/263) for open surgery and 69 per cent (55/80) for keyhole surgery. The main reason that the success rate reversed is because the choice of surgery depends on the diameter of the stones.

Although randomised-controlled trials are more scientifically rigorous than observational studies, it is not always feasible to do randomised trials for a variety of reasons. The section on alcohol and car accidents above is one example. Another 'tongue-in-cheek' example is the use of a parachute as an intervention to prevent death or serious injury after jumping out of an aeroplane.[6] The authors point out in a highly articulate and extremely witty way that we must not always jump (pun not intended) to conclusions when assessing evidence of interventions resulting from observational data. In the discussion, the authors make the following point related to selection bias: 'individuals jumping from aircraft without the help of a parachute are likely to have a high prevalence of pre-existing psychiatric morbidity'. If you have not come across this article yet, I strongly recommend that you get a copy.

I hope you enjoyed reading this article and it served to educate rather than to confuse. I would like to leave you with this famous Aaron Levenstein quote: 'Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital'. ■

> **"Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital"**

REFERENCES

1. **http://en.wikipedia.org/wiki/List_of_paradoxes**
2. **Goldacre B.** *Bad Science.* London: Fourth Estate Ltd, 2008.
3. **Gill J.** *Bayesian Methods.* Boca Raton, FL, USA: CRC Press, 2002.
4. **Simpson EH.** The interpretation of interaction in contingency tables. *J R Stat Soc B* 1951; 13: 238–41.
5. **Julious SA, Mullee MA.** Confounding and Simpson's paradox. *Brit Med J* 1994; 309(6967): 1480–1.
6. **Smith GC, Pell JP.** Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Brit Med J* 2003; 327(7429): 1459–61.