Coincidences: the truth is out there

KEYWORDS:
Teaching;
Probability;
Birthday paradox;
Coincidences.

Robert Matthews
University of Northern Colorado, USA.
Fiona Stones
Granada TV, London, England.

Summary
We give a simple way of demonstrating that coincidences really are “out there”, as probability theory predicts, if we take the trouble to look

◆INTRODUCTION ◆

COINCIDENCES often surprise us, suddenly springing up to reveal an unexpected connection between people or things. Sometimes they seem so outlandish as to demand a “supernatural” explanation. Yet anyone familiar with probability theory knows how this notoriously counter-intuitive branch of mathematics can spring big surprises on us. One of the most famous and relevant examples is the so-called Birthday Paradox, which states that in a random gathering of just 23 people, there are 50:50 odds that at least two of those present have the same birthday. Many people find this result very surprising: a recent survey of university students found a median value for the estimated size of gathering needed of *385* (Matthews & Blackmore (1995)). So large a gathering is, of course, *guaranteed* to contain at least one coincident birthday, suggesting that the probabilistic aspects of the paradox evade many people. The same study also showed that people tend to grossly overestimate the size of gathering needed for other types of coincidence.

Part of the explanation for this general lack of insight into the probability of coincidences is that most of us do not go “looking” for coincidences: they “find” us. If instead we made a point of demanding the birthdays of everyone at every gathering we attend, we would soon discover that coincident birthdays are indeed relatively common. We would also get a better understanding of why: firstly, that we are not demanding a coincidence between *specific* people or *specific* birthdays, but just *any* people and *any* birthday; and secondly, that the key factor is not the number of people at the gathering, *N*, but the very much larger number of possible pairings of people with which to get a match, $N(N—1)/2$ (= 253 in the case of N = 23). A more quantitative explanation of the Birthday Paradox can, of course, be given using probability theory. There is, however, no substitute for real-life evidence, and in what follows we outline a simple and appealing demonstration that coincidences really are “out there” and they follow the predictions of probability theory.

◆COINCIDENCES◆ IN FOOTBALL MATCHES

The key source of perplexity with the Birthday Paradox is the low number of people needed to give decent odds of finding at least one coincident birthday. A football match provides an ideal test-bed for this assertion: it has 23 people on the pitch (11 players per side, plus the referee), and it seems reasonable to assume their birthdays are randomly distributed over the year (a point we return to later). If the Birthday Paradox is correct, then in a sample of F fixtures, we expect about $0.5F$ to contain at least one pair of players sharing the same birthday. However, probability theory allows us to predict several other types of coincidence we should also expect to observe. To see this, we can model the distribution of players’ birthdays among the days of the year as a balls-in-urns model, with 23 “balls” being distributed among 365 “urns”. A coincidence is then characterised by having at least one urn containing two or more balls, a situation that can be visualised via an “occupancy diagram’’:

$$\begin{array}{ccc} [2] & [1,...,1] & [0, ... ,0] \\ (1) & (21) & (343) \end{array}$$

where the’ numbers in square brackets show the occupancy of each urn, and the numbers in parentheses represent the numbers of urns with these levels of occupancy. This diagram represents the case of *precisely* two of 23 people sharing the same birthday. Calculating the probability of such an arrangement is then a three-stage process: (i) calculation of the number of ways of arranging the urns, *U;* (ii) calculation of the number of ways of arranging the balls within those urns, *B;* (iii) multiplying *U* by *B* and dividing by the total number of ways of distributing 23 balls among 365 urns, i.e.365^{23} Both (i) and (ii) are given by the standard result that the number of ways of dividing a population of N elements into k sub-groups, of which the first contains r_1 elements, the next r_2 elements and so on, is $N!/(r_1!r_2!...r_k!)$. We then obtain the following results:

(1) Probability of at least one coincident birthday

This is 1 - P(no coincident birthday), where the lack of a coincident birthday leads to an occupancy diagram of

[1,...,1]	[0, ..., 0]
(23)	(342)

U is 365!/(23!342!), while B is $23!/(1!)^{23}(0!)^{342}$ = 23! and thus P(no coincident birthday) = 365^{-23} x 365!/342! = 0.493, so that P(>1 coincident birthday) = 0.507, from which the original Birthday Paradox follows.

(2) Probability of precisely one coincident birthday

The occupancy diagram was given earlier, and leads to U= 365!/(21!343!), and B = 23!/2! and so P(1 coincident birthday) = 365^{-23} x U x B = 0.363.

(3) Probability of precisely two coincident birthdays

For two pairs of participants to share coincident birthdays, the occupancy diagram is

[2, 2]	[1,...1,]	[0, ... ,0]
(2)	(19)	(344)

so that U= 365!/(2! 19! 344!), and B = $23!/(2!)^2$ and P(2 coincident birthdays) = 0.111

(4) Probability of precisely three coincident birthdays

For three pairs of participants to share coincident birthdays, the occupancy diagram is

[2, 2, 2]	[1,...,1]	[0, ...,0]
(3)	(17)	(345)

so that U= 365!/(3! 17! 345!),and B = $23!/(2!)^3$ and
P(3 coincident birthdays) = 0.018

(5) Probability of one set of triply-coincident birthdays

For three participants to share the same birthday, the occupancy diagram is

[3]	[1 1]	[0 0]
(1)	(20)	(344)

so that U=365!/(1! 20! 344!), and B = 23!/3! and P(1 triply-coincident birthday) = 0.007

(6) Probability of birthday on day of fixture

To demonstrate the impact of being *specific* about the day for which a coincidence is required, we also include the probability that at least one person among N playing on a specific day will be celebrating their birthday. This is 1 - $(364/365)^N$ = 0.061 for N = 23; for P(birthday on specific day) to be 0.5 requires N around 256.

Having shown how to calculate probabilities of various types of coincidences occurring in a football fixture, let us now put them to the test.

◆ANALYSIS ◆ OF FOOTBALL FIXTURES

To find out if the various birthday coincidences do occur at the rate predicted above, we need a sample of football fixtures, and the dates of birth of all the players and referees. For our sample, we chose the ten Premier Division fixtures played on 19 April 1997, which at kick-off involved a total of 220 players and 10 referees. We obtained the dates of birth of players using Rollin (1996) plus some club data, while the referees' data came from the Football Association (we note that it is not necessary to use referees: the dates of birth of the first substitute played could be used instead). By cross-checking the various dates of birth in all 10 fixtures, we obtained the following results:

Table **1.** Coincident birthdays in Premiership fixtures on 19 April 1997.

Fixture (team, date)	Coincident birthdays
Arsenal v. Blackburn	No coincidences
Aston Villa v. Tottenham	Eliogu (AV; 3.11.72) Yorke (AV; 3.11.71)
Chelsea v. Leicester City	Petrescu (C; 22.12.67) and Morris (C; 22.12.78) Hughes (C; 1.11.63) and Elliott (LC; 1.11.68)
Liverpool v. Manchester Utd	James (L; 1.8.70) and Wright (L; 1.8.63) Butt (M; 21.1.75) and P Neville (M; 21.1.77)
Middlesbrough v. Sunderland	Johnston (5; 14.12.73) and Waddle (5; 14.12.60)
Newcastle v. Derby	No coincidences
Nottingham Forest v. Leeds	Martyn (Le; 11.8.66) and Halle (Le; 11.8.65)
Sheffield Wed v. Wimbledon	No coincidences
Southampton v. Coventry	Benali (So; 30.12.68) and Whelan (Co; 30.12.74)
West Ham v. Everton	No coincidences

We can now compare these results to the number of coincidences of various types expected to occur among 10 fixtures using the probabilities calculated in the previous section. The results are as follows (the expected frequencies are rounded to integers):

Table 2. Comparison of expected and observed number of coincidences in 10 fixtures.

Type of coincidence	Expected	Observed
No coincidence seen	5	4
At least one coincident birthday	5	6
Exactly one coincident birthday	4	4
Exactly two coincident birthdays	1	2
Exactly three coincident birthdays	0	0
Exactly one triply-coincident birthday	0	0
>1 participant with birthday on 19.4.97	0 - 1*	0

* Based on $1 - (364/365)^{230} = 0.47$

Table 2 shows impressive agreement between the predictions of probability theory and the observed number of coincidences. In particular, it confirms the theoretical prediction that the less specific a coincidence is, the more likely it is to occur: getting *any* two players to share *some* birthday proved possible in four out of the 10 fixtures, but not one of all 230 participants had a birthday on the *specific* day of the match.

◆ THE ‘NEAR MISS” EFFECT ◆

As we have seen, coincidences tend to be considerably more likely than we might think. They become more likely still if we allow a little latitude into our definition of what constitutes a coincidence - for example allowing birthdays separated by no more than *r* days of each other to constitute a “hit”. As before, we can model this “nearmiss effect” using a balls-in-urns model; the argument is somewhat more involved (see for example Naus (1968)) and leads to

$$P(>2 \text{ birthdays separated by } < r \text{ days}) = 1 - [(364 - rN)!365^{1-N} / (365 - (r+1)N)!]$$

Thus, for exactly coincident birthdays we have $r = 0$, while for birthdays either on the same day or on adjacent days we have $r = 1$. To compare these theoretical values with the reality of our football matches, we set $N = 23$, leading to a near-miss probability of 0.888 for $r = 1$: that is, we expect about 9 of the 10 fixtures to feature participants whose birthdays are within a day of each other. In fact, all 10 of the matches have at least two players with birthdays within a day of each other - again, impressive agreement with the predictions of probability theory.

The ability of the “near-miss effect” to boost considerably the chances of coincidences can be seen even at the level of each team. Setting $N = 11$, we find that the probabilities of individual teams having at least two birthdays separated by no more than 0, 1, 2, 3, 4 and 9 days are 0.141, 0.371, 0.543, 0.672, 0.767 and 0.948 respectively. Table 3 compares this with observation:

Table 3. Comparison of expected and observed number of “near-miss” coincidences among 20 teams.

Type of coincidence	Expected	Observed
At least two coincident birthdays	3	6
At least two birthdays < 1 day apart	7	13
At least two birthdays < 2 days apart	11	17
At least two birthdays < 3 days apart	13	18
At least two birthdays <4 days apart	15	18
At least two birthdays < 9 days apart	19	20
>1 participant birthday on (18-20)/4/97	1	1

Once again, the overall agreement between theory and observation is impressive: as before, we see that as the “window of opportunity” given to the near-miss effect is widened, the number of coincidences increases. It is worth noting that the biggest increase comes from allowing birthdays that fall on adjacent days also to count as “hits”: this small concession doubles the number of coincidences. It is also worth pointing out that - as with our earlier comparisons of theory with reality - the deviations between theory and observation tend to favour the exi~tence of more coincidences. This is a reflection of the fact that there is a significant preponderance of players’ birthdays in November and December, and deviations away from a uniform distribution of birthdays always tend to boost still further the number of observed coincidences.

◆CONCLUSION ◆

We have shown that football fixtures provide a simple and convenient way of investigating the prevalence of coincidences. The raw data are of a familiar type, are easy to obtain from published sources, and motivate the use of simple combinatorics in making predictions about what should be observed. Our own previous research suggests many people will be very surprised by the results.

Acknowledgements

RM thanks Simon Singh for inspiring him to use football matches as a way of probing coincidences, John Haigh and an anonymous referee for very helpful comments, and Ms J A Steam of the Football Association for the dates of birth of referees.

References

Matthews, R.A.J., Blackmore, SI. (1995). Why are coincidences so impressive? *Perceptual and Motor Skills*, **80,** 1121-1122.

Naus, I. I. (1968). An extension of the Birthday Problem, *The American Statistician*, 22, 27-29.

Rollin, J. (1996). *Guinness Soccer Who’s Who,* **10th Edition** (Guinness Publishing, Enfield).