

# Historias de Matemáticas

## A probabilistic approach to coincidences: the birthday paradox

## Un enfoque probabilístico para las coincidencias: la paradoja del cumpleaños

Carla Santos y Cristina Dias

Revista de Investigación



Volumen V, Número 2, pp. 055–060, ISSN 2174-0410

Recepción: 15 May'15; Aceptación: 18 Sep'15

1 de octubre de 2015

### Resumen

Ante la ocurrencia simultánea de eventos que consideramos muy poco probables, que llamamos coincidencia, todos quedamos sorprendidos. Sin embargo, la búsqueda de explicación para la ocurrencia de una coincidencia puede hacerse por diferentes enfoques. Desde el punto de vista de Diaconis y Mosteller, las coincidencias no son tan raras como creemos. La paradoja del cumpleaños ilustra la idea de que algo muy improbable desde el punto de vista individual, puede ocurrir un considerable número de veces, en general. Para ilustrar la validez de esta paradoja usaremos los cumpleaños de las listas de los escuadrones oficiales de la Copa del Mundo FIFA 2014.

**Palabras Clave:** Coincidencias, paradoja del cumpleaños, probabilidad, Copa del Mundo de la FIFA de 2014.

### Abstract

Considering the simultaneous occurrence of events that we consider highly improbable, the so called coincidences, we all get surprised. However the search for the explanation of this occurrence can be made by different approaches. From the standpoint of Diaconis and Mosteller, coincidences are not as unusual as we think. The birthday paradox illustrates the idea that something highly improbable from the individual point of view may, however, occur a considerable amount of times in general. To illustrate the validity of this paradox we will use the birthdays from FIFA's official squad lists of 2014 World Cup.

**Keywords:** Coincidences, Birthday paradox, probability, 2014 FIFA World Cup.

## 1. Introduction

The perception that the simultaneous occurrence of certain events is almost impossible makes it be seen as something extraordinary, what we call coincidence. Before the surprise that the occurrence of this phenomenon creates and the difficulty in understanding it, there is a need to provide an explanation for it. This explanation may lie in mystical arguments, associated to some conspiracy theory, in synchronicity theory or using other approaches.

Although, for each one of us individually, a coincidence is seen as something extraordinary, considering a large number of events, extremely improbable coincidences will have high probability to occur (Diaconis and Mosteller, 1999). The probability theory allows us to determine the probability of these coincidences to occur.

If coincidences of birthday dates are much more common than we would think, then probably many of those other coincidences we are faced up, are much easier to happen than we suppose and aren't not that much extraordinary.

## 2. Coincidences

The small probability with which certain events happen simultaneously makes this occurrence to be seen as something extraordinary. We call it coincidence. Diaconis and Mosteller (1989) define coincidence as "a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection".

Nowadays, as always, the fascination for coincidences is high. Although there is not an accepted universal explanation, numerous scientists and researchers have suggested several theories.

Carl Jung, XX century, psychiatrist, tried to discover the reason for the existence of coincidences having created the synchronicity theory, in which the coincidence of facts in space and time was seen as something more than purely random, proposing the existence of a link between psychic and physical events. Despite the difficulty to prove this link, Jung's theory still has followers today. For example, Chopra, D. (2003), claim that "coincidences are not accidents but signals from the universe which can guide us toward our true destiny".

Amongst the ones that adopted a more skeptical vision, the attribution of meaning to coincidences is totally due to human nature itself. One of the explanations presented to the great relevance we attribute to coincidences is the apophenia, i.e., predisposal of our mind to try to identify connections and patterns in random data or meaningless. Another possible explanation, named egocentric bias, was given by Falk (1989), showing that the perception that something extraordinary occurred is highlighted when there is personal involvement in that event, having the predisposition to consider it subjectively less likely to occur than if it had happened to someone else.

In reality, events, that from an individual point of view are very slightly probable, occur with high frequency when a large number of people is involved. As Diaconis and Mosteller (1989) state in their law of truly large numbers:

*"With a large enough sample, any outrageous thing is likely to happen."*

Let's suppose that an incredible coincidence happens per day to one person in a million. In a country like Portugal, with about 10.5 million people, in a year there will occur 3832 incredible coincidences. In the whole world, considering a population of 7 billion people, there will occur, annually, over two million and a half incredible coincidences.

The fact that we don't acknowledge the high number of opportunities for coincidences that day to day life provides, together with the incapacity of estimating the probability for the occurrence of these events, leads us to underestimate the probability for the occurrence of coincidences.

### 3. The Birthday Paradox

In contrast to what happens in some branches of Mathematics, the Probability Theory is fraught with problems that are easily understood by everyone, due to its close connection to everyday phenomena. This deceptive simplicity makes that, when facing a problem involving probabilities associated with a familiar phenomenon, the proposed solutions are supported only by common sense and intuition. However, these problems often mask a complex reasoning leading to counterintuitive results.

One of the most famous problems whose result has a difficult intuitive acceptance is the "birthday paradox". Although the "birthday paradox" is not a true mathematical paradox, often takes this designation (see Székely, 1986) because it gives a surprising response, contrary to what common sense would indicate.

Since it have been proposed by Richard von Mises, in 1939, the birthday paradox has occurred frequently in the literature under different perspectives considering, for example, non-uniform birth frequencies (see Joag-Dev and Proschan, 1992; Mase, 1992; Henze, 1998; Camarri and Pitman, 2000) or generalizations (see Székely, 1986; Nishimura and Sibuya, 1988; Flajolet et al., 1992; Polley, 2005; McDonald, 2008). Among the different applications of the birthday paradox we can detach cryptography (e.g. Suzuki et al. 2008; Galbraith and Holmes, 2010) and forensic sciences (e.g. Weir, 2007; Kaye, 2012; Obasogie, O.K., 2013).

In the statement of this problem is asked:

*"How many people should be in a room so that the probability of two of them celebrate their birthday on the same day exceeds 50%?"*

This is the simplest version, based on the assumptions that a year has 365 days (ignoring the existence of leap years), birthdays are independent from person to person and that the 365 possible birthdays are equally likely. (see Feller, 1968).

As there are 365 possible dates for birthdays, we have the (correct) perception, that is very unlikely to find someone with the same birthday as us. Whereas the distribution of birthdays is uniform, randomly choosing a person, the probability of that person celebrating the birthday on the same day that another person is  $\frac{1}{365}$ , this is, there is only one day, among 365 days of the year, in which the birthday dates may coincide. Therefore, the probability of two people have the same birthday is extremely low,  $\frac{1}{365} = 0,0027 = 0,27\%$ . Even if one person compare his birthday with 20 people, the probability of match remains relatively low, about 5%.

As a matter of fact, the question is not about the probability of a certain person of the group having the same birthday date than one other person picked randomly, but the probability of some of the persons of the group having a common birthday date with any other person of that group. The detail that makes all the difference in this problem is the fact that in a group of people each one of them can check with each one of the others if their birthdays coincide.

## 4. Solving the Birthday Paradox

Trying to find, in a group of  $k$  persons, at least one person with the same birthday as another can be seen as a case of sampling without replacement, Parzen (1960).

Let  $p_k$  be the probability that, in a group of  $k$  persons, at least one person have the same birthday date than another, and  $q_k = 1 - p_k$  the probability that all  $k$  persons have different birthdays.

Let us suppose the group has only two persons, this is,  $k = 2$ . The first person can have his birthday on any of the 365 days of the year and the second one can have his birthday on any of the remaining 364 days. The probability that these two persons do not have the same birthday is

$$q_2 = 1 - \frac{1}{365} = \frac{364}{365} = 0,997$$

If we add one more person to the group, this third person cannot share the birthday with any of the other two, in order to, all three, have different birthdays. For the third person there are 363 available days, so the probability that three persons do not have the same birthday is

$$q_3 = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) = \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) = 0,992$$

Generalizing this for a group of  $k$  persons, the probability that all  $k$  persons have different birthdays is

$$\begin{aligned} q_k &= \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right) = \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{365-k+1}{365}\right) = \\ &= \frac{364!}{365^{k-1}(365-k)!} = \frac{365!}{365^k(365-k)!} \end{aligned}$$

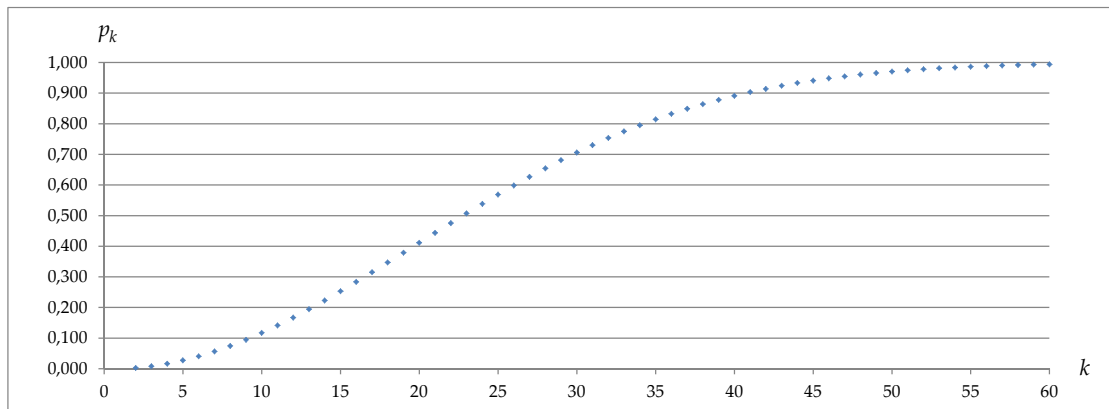
To have  $p_k$  larger than 50%,  $q_k$  must be less than 50%. Then, we have to find the value of  $k$ , for which the probability,  $q_k$ , that all  $k$  persons have different birthdays, is less than 50%. Using Microsoft Excel (or any other similar program), to calculate the corresponding probability for  $k = 2, 3, 4, \dots$  onwards until you can get a value for  $p_k$  larger than 0,5, we find  $k = 23$  (see Table 1).

Table 1.  $q_k$  and  $p_k$  values.

$k$	$q_k$	$p_k$	$k$	$q_k$	$p_k$
2	0,997	0,003	21	0,556	0,444
3	0,992	0,008	22	0,524	0,476
4	0,984	0,016	23	0,493	0,507
5	0,973	0,027	30	0,294	0,706
10	0,883	0,117	50	0,030	0,97
20	0,589	0,411	57	0,010	0,99

As we can see on Table 1 or in Graph 1, the first value for which the probability  $p_k$  is larger than 0,5 is 23.

So, if there are 23 persons in a room, the probability of, at least, two of them share the birthday is a little more than 50%. With 50 persons this probability increases to 97%, and with "only" 57 persons the probability is 99%.

Graph 1.  $p_k$  values for  $2 \leq k \leq 60$ .

## 5. A curious illustration: The birthday paradox in 2014 FIFA's world cup

To test the birthday paradox we used the birthdays from FIFA's official squad lists of 2014 World Cup.

In this World Cup, 32 squads were in competition, each team with 23 players. The birthday paradox states that a group of 23 people the probability of two of them celebrate their birthday on the same day is approximately 50%. Then, we expect that about half of these teams have at least a couple of players to celebrate their birthday on the same day.

Based on the biographical data of the players available on the FIFA website, we found that there are 11 squads (Australia, United States of America, Cameroon, Bosnia and Herzegovina, Russia, Nigeria, Spain, Colombia, Netherlands, Brazil and Honduras) with one pair of players celebrating the birthday on the same day, and 5 teams (Iran, France, Argentina, South Korea and Switzerland) with two pairs of players with the same birthday.

Since 16 of the 32 teams have shared birthdays, the Birthday Paradox is confirmed!

## References

- [1] CAMARRI, M. and PITMAN, J., "Limit distributions and random trees derived from the birthday problem with unequal probabilities", *Electron. J. Probab.* 5, 1–18, 2000.
- [2] CHOPRA, D., *The Spontaneous Fulfillment of Desire* Harmony Books, New York, 2003.
- [3] COPPERSMITH, D., *Another birthday attack*, *Advances in Cryptology, Proc. of Crypto'85*, LNCS, vol. 218, Springer-Verlag pp. 14–17, 1986.
- [4] DIACONIS, P. and MOSTELLER, F., "Methods of Studying Coincidences", *Journal of the American Statistical Association*, vol 84, No 408, 1989.
- [5] FALK, R., "The Judgment of Coincidences: Mine Versus Yours", *Amer. J. Psych.* 102, 477–493, 1989.
- [6] FELLER, W., *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed., John Wiley, New York, 1968.

- [7] FLAJOLET, P., GARDY, D., and THIMONIER, L., "Birthday paradox, coupon collectors, caching algorithms, and self-organizing search", *Discrete Applied Mathematics* 39, 1992.
- [8] GALBRAITH, S. D. and HOLMES, M. "A non-uniform birthday problem with applications to discrete logarithm", *IACR Cryptology ePrint Archive* 2010: 616, 2010.
- [9] JOAG-DEV, K. and PROSCHAN, F. "Birthday problem with unlike probabilities", *American Mathematical monthly* 99: 10–12, 1992.
- [10] KAYE, D. H., "Beyond Uniqueness: The Birthday Paradox, Source Attribution, and Individualization in Forensic Science Testimony Law", *Probability & Risk*, 2012.
- [11] MASE, "Approximations to the birthday problem with unequal occurrence probabilities and their application to the surname problem in Japan", *Annals of the Institute of Statistical Mathematics*, 44, 479–499, 1992.
- [12] MATTHEWS, R. A. J. and BLACKMORE, S. I., "Why are coincidences so impressive?", *Perceptual and Motor Skills*, 80,
- [13] MCDONALD, M. P. and JUSTIN LEVITT, J., "Seeing double voting: An extension of the birthday problem". *Election Law Journal*, 7(2): 111–122, 2008.
- [14] MERKUR, D., *Mystical Moments and Unitive Thinking*. State University of New York Press, Albany, NY, 1999.
- [15] NISHIMURA, K. and SIBUYA, M., "Occupancy with two types of balls", *Ann. Inst. Statist. Math.*, 40, 77–91, 1988.
- [16] OBASOGIE, O. K., "High-Tech, High-Risk Forensics", *N.Y. Times*, July 25, 2013.
- [17] PARZEN, E., *Modern Probability Theory and Its Applications*, John Wiley & Sons, 1960.
- [18] PIPES, D., *Conspiracy: How the Paranoid Style Flourishes and Where It Comes From*, New York: Touchstone, 1997.
- [19] POLLEY, W. J., "A Revolving Door Birthday Problem", *UMAP Journal*, Vol. 26 Issue 4, p. 413, 2005.
- [20] SU, C. and SRIHARI, S. N., "Generative Models and Probability Evaluation for Forensic Evidence", P. Wang (ed.), *Pattern Recognition, Machine Intelligence and Biometrics*, Springer, 2011.
- [21] SUZUKI, K., TONIEN, D., KUROSAWA, K. and TOYOTA, K., "Birthday paradox for multi-collisions". *E91-A(1):39–45*, 2008.
- [22] SZÉKELY, G. J., "Paradoxes in Probability Theory and Mathematical Statistics". Akadémiai Kiadó, Budapest, 1986.
- [23] WEIR, B.S., "The rarity of DNA profiles". *Annals of Applied Statistics*; 1: 358–370, 2007.

### Sobre las autoras:

*Nombre:* Carla Santos

*Correo electrónico:* carla.santos@ipbeja.pt

*Institución:* Instituto Politécnico de Beja. Portugal.

*Nombre:* Cristina Dias

*Correo electrónico:* cpsilvadias@gmail.com

*Institución:* Instituto Politécnico de Beja. Portugal.