SMSW'14 Statistics and Mathematical Sciences Workshop
UBI, July 7 - 2014 in Honour of Professor João Tiago Mexia

# Coincidences are more likely than you think: The birthday paradox

**Carla Santos** [1] **and Cristina Dias** [2]

*[1]Polytechnical Institute of Beja and CMA - Center of Mathematics and its Applications*
*[2]Polytechnical Institute of Portalegre and CMA - Center of Mathematics and its Applications*

Funded by
PEst-OE/MAT/UI0297/2014

FCT
Fundação para a Ciência e a Tecnologia

# Outline

1. **Introduction**
2. **Coincidences**
3. **The birthday paradox**
4. **The birthday paradox in 2014 Football WorldCup**

# Introduction

The perception that simultaneous occurrence of certain events is practically impossible makes it be seen as something extraordinary, that we call coincidence.

# Coincidences

Diaconis & Mosteller (1989) define coincidence as

"a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection".

# Coincidences

Although there is no universally accepted explanation for coincidences, various scientists and researchers have proposed several theories.

Carl Jung, XX century psycanalyst, tried to discover the reason for the existence of coincidences in his **Synchronicity Theory** where he proposes the existence of a link between psychic and physical events.

For others, with a more skeptical vision, the attribution of meaning to coincidences is totally due to human nature itself:

## Apophenia

predisposal of our mind to try to identify connections and patterns in random or meaningless data .

## Egocentric bias

the highlighting of the perception that something extraordinary occurred when there is personal involvement in that event . (Falk,1989)

# Coincidences

Diaconis and Mosteller (1989, p. 859) say that the relevant principle to use when reasoning about coincidences is an idea they term as

**Law of Truly Large Numbers**

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

# Coincidences

We underestimate the probability
for the occurrence of coincidences

We don’t acknowledge the high number of opportunities for coincidences that day to day life provides

We are incapable of estimate the probability for the occurrence of these events

# Coincidences

**Let's suppose that an incredible coincidence happens per day to one person in a million.**

# Coincidences

In a country like Portugal, with 10,5 million people, in a year, there will occur 3832 incredible coincidences.

# Coincidences

In the whole world, considering a population of 7 billion people, there will occur over 2,5 million incredible coincidences, in a year.

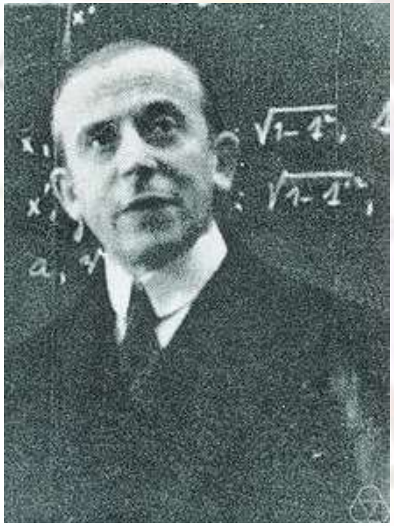
# Birthday paradox

A good way to illustrate the idea that something highly improbable from the individual point of view may, however, occurs a considerable amount of times in general, is the **Birthday Paradox**[1]**.**

[1] Althought the Birthday Paradox is not a real paradox ( a statement or a concept that seems to be self-contradictory) it takes this name because it origins a surprising answer that is against the common sense (Székely, 1986).

# Birthday paradox

Since it have been proposed by Richard von Mises, in 1939, the birthday paradox has occurred frequently in the literature under different perspectives, for example, considering non-uniform birth frequencies (see Mase, 1992; Camarri and Pitman, 2000) and generalizations (see Székely, 1986; Polley, 2005; McDonald,2008).



# Birthday paradox

Applications of the Birthday paradox

- Cryptography (e.g. Coppersmith ,1986; Galbraith and Holmes,2010)
- Foresic Sciences (e.g. Su C. and Srihari S. N., 2011;).


Fig. 13. Three specific fingerprints (from the same finger) used to calculate probabilities: (a) good quality full print $F_1$, (b) low quality full print $F_2$, and (c) partial print $F_3$. **In Su C. and Srihari S. N., 2011**

# Birthday paradox

The simplest and more popular formulation of the birthday paradox asks:
(see e.g. Feller,1968; Berresford,1980)

How many people you need to have in a room so that there is a better-than-even chance that two of them will share the same birthday?

This version is based on the assumptions that:

- a year has 365 days (ignoring the existence of leap years)
- birthdays are independent from person to person
- the 365 possible birthdays are equally likely.

# Birthday paradox

How many people you need to have in a room so that there is a better-than-even chance that two of them will share the same birthday?

The answer to Birthday Paradox question is surprisingly low,

23

# Birthday paradox

The birthday paradox is counter-intuitive because we tend to view the problem from our own individual perspective.

Considering there are 365 days in a year, we consider extremely unlikely to find someone who shares our birthday date.

In fact, the probability of two persons have their birthday on the same day is extremely low, $1/365 = 0.0027 = 0.27\ \%$.

# Birthday paradox

✘ But the question is not about the probability of a certain person of the group having the same birthday date than one other person picked at random!

✓ In a group of people, each one of them can check with each one of the others if their birthdays match!

# Birthday paradox

**The usual "exact" calculation**

Trying to found at least one person with the same birthday that one other in a group of $k$ persons, can be considered a case of sampling with replacement (Parzen,1960)

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

# Birthday paradox

**The usual "exact" calculation**

If the group has only 2 persons:

- The first person can have his birthday on any of the 365 days of the year;
- The second person has 364 available dates for his birthday.

The probability of they do not share their birthday is then:

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

# Birthday paradox

**The usual "exact" calculation**

Lets add one more person to the group:

In order to all of them have different birthdays, the 3rd person´s birthday can not match with any of the others birthday.
The probability of the 3 persons celebrate their birthdays in different dates is:

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

# Birthday paradox

**The usual "exact" calculation**

In a group of $k$ persons, the probability of all of them celebrate their birthdays in different dates is:

$$q_k = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right) =$$

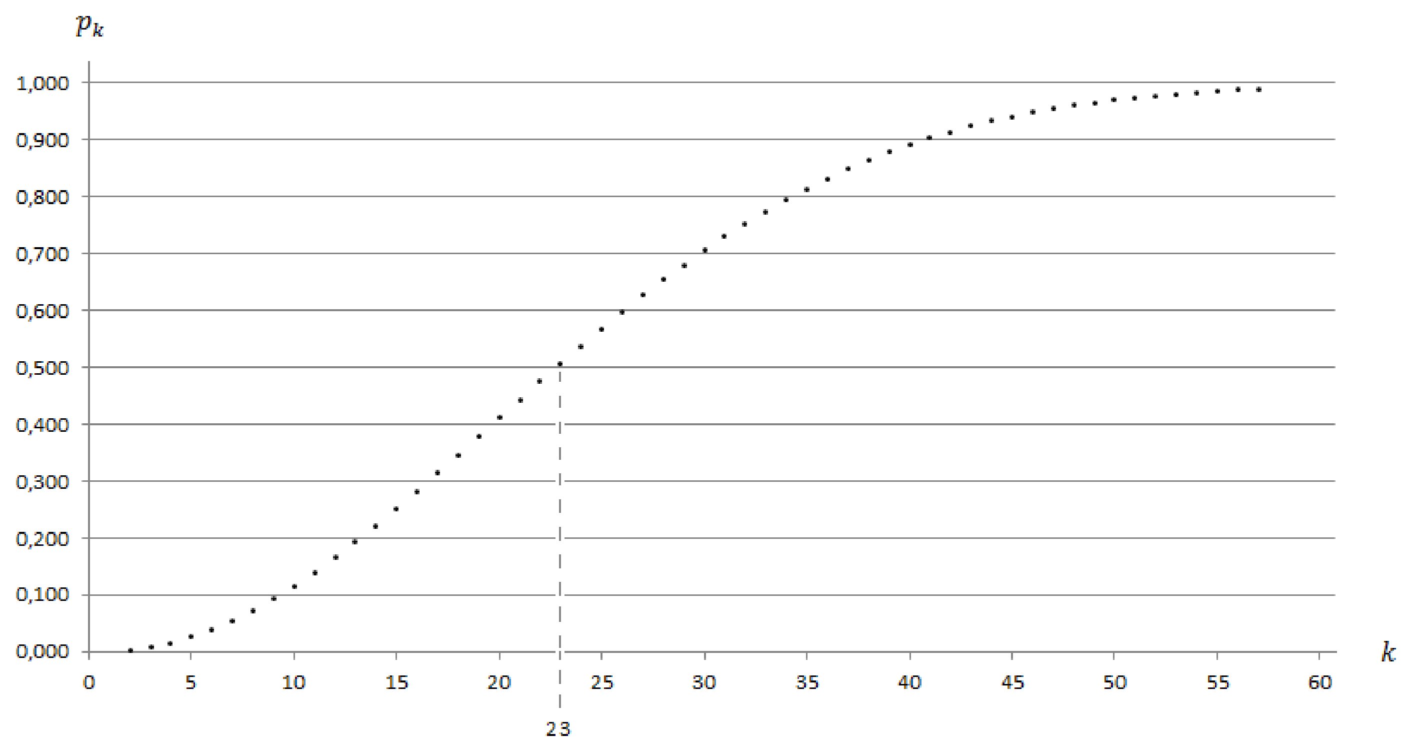
$$= \left(\frac{364}{365}\right)\left(\frac{363}{365}\right) \cdots \left(\frac{365-k+1}{365}\right) = \frac{364!}{365^{k-1}(365-k)!} =$$

$$= \frac{365!}{365^{k}(365-k)!}$$

# Birthday paradox

**The usual "exact" calculation**

The first value of $k$ for which the probability $p_k$ is more than 50% is $k$ =23.



# Birthday paradox

## The alternative "exact" calculation

Consider the number of comparisons among all the elements of the group.

Since each one of the persons have to check with each one of the others if their birthdays match, in a group of $k$ persons, the total number of comparisons will be

$$i = \binom{k}{2} \quad ,$$

the number of possible combinations (without replacement) of $k$ elements taken 2 at a time.

# Birthday paradox

**The alternative "exact" calculation**

In each one of the comparisons, the probability of matching of birthday dates is $\frac{1}{365}$, so the probability that there is no match in $i$ comparisons is

$$\left(1-\frac{1}{365}\right)^i=\left(\frac{364}{365}\right)^i,$$

and the probability that there is at least one match is

$$1-\left(\frac{364}{365}\right)^i.$$

# Birthday paradox

**The alternative “exact” calculation**

The lowest number of comparisons that have to be made in order to have a probability, of two persons have the same birthday, greater than 50%, is 253.

$$1-\left(\frac{364}{365}\right)^{i} \geq 0{,}5$$

$$i \geq 252{,}65$$

# Birthday paradox

**The alternative "exact" calculation**

To have 253 comparisons in a group of $k$ persons,

$$253 = \binom{k}{2} \Leftrightarrow k^2 - k - 506 = 0 \, .$$

Then, the group must have 23 persons.

# Birthday paradox

**The Poisson approximation**

Let $X$ be a random variable, representing the number of birthday's matches, among $k$ persons.

$$X \sim B(i\,,p)$$

where $i = \binom{k}{2}$ and $p = \frac{1}{365}$.

# Birthday paradox

**The Poisson approximation**

Since $p < 0{,}1$ and $ip > 5$ , Arrantia (1990) proposes to use the Poisson distribution with $\lambda = \binom{k}{2}\frac{1}{365}$.

Then, the probability of having at least one birthday's match is:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\frac{k(k-1)}{730}}$$

Solving the inequation $1 - e^{-\frac{k(k-1)}{730}} \geq$ 0,5, we find, once more, $k$ =23.

# Birthday paradox

## In Football Worldcup 2014

To test the birthday paradox I used the birthdays from FIFA's official squad lists of 2014 World Cup.
In this World Cup, 32 teams were in competition and each team has 23 players.

| | | | |
|---|---|---|---|
| Algeria | Argentina | Ghana | Greece |
| Australia | Belgium | Honduras | Iran |
| Bosnia and Herzegovina | Brazil | Italy | Japan |
| Cameroon | Chile | Korea Republic | Mexico |
| Colombia | Costa Rica | Netherlands | Nigeria |
| Côte d'Ivoire | Croatia | Portugal | Russia |
| Ecuador | England | Spain | Switzerland |
| France | Germany | Uruguay | USA |

# Birthday paradox

**In Football Worldcup 2014**

Based on the biographical data of the players available on the FIFA website it turns out that:

| One pair with the same birthday | Two pairs with the same birthday |
|---|---|
| Australia, United States of America, Cameroon, Bosnia and Herzegovina, Russia, Nigeria, Spain, Colombia, Netherlands, Brazil and Honduras | Iran, France, Argentina, South Korea and Switzerland |

There are 11 teams with one pair of players that celebrate the birthday on the same day, and 5 teams with two pairs of players with the same birthday.

Since 50% of the teams have shared birthdays, the Birthday Paradox is confirmed!

# Birthday paradox

**References:**


- Arratia, R. , Goldstein, L. and Gordon, L. (1990) Poisson Approximation and the Chen-Stein method. Statist. Sci. 5, 403-434.
- Berresford, G. C. , 1980.The Uniformity Assumption in the Birthday Problem, Mathematics Magazine, Vol. 53, No. 5, pp. 286-288
- Coppersmith, D. , 1986, Another birthday attack. Advances in Cryptology, Proc. of Crypto'85, LNCS, voi. 218, Springer- Verlag pp. 14-17.
- Diaconis, P. , Mosteller, F. , 1989. Methods of Studying Coincidences, Journal of the American Statistical Association, vol 84, No 408.
- Falk, R. (1989). The Judgment of Coincidences: Mine Versus Yours. Amer.J. Psych. 102, 477-493.
- Feller, W. (1968). An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., John Wiley, New York.
- Galbraith, S.D. , Holmes, M. (2010). A non-uniform birthday problem with applications to discrete logarithms. IACR Cryptology ePrint Archive 2010: 616
- Merkur, D. (1999). Mystical Moments and Unitive Thinking. State University of New York Press, Albany, NY.
- Parzen, E. , 1960. Modern Probability Theory and Its Applications, John Wiley & Sons.
- Pipes, D. (1997). Conspiracy: How the Paranoid Style Flourishes and Where It Comes From. New York: Touchstone
- Su C. and Srihari S. N., 2011, Generative Models and Probability Evaluation for Forensic Evidence," in P. Wang (ed.), Pattern Recognition, Machine Intelligence and Biometrics, Springer.
- Székely, G. J. ,1986. Paradoxes in Probability Theory and Mathematical Statistics. Akadémiai Kiado, Budapest.