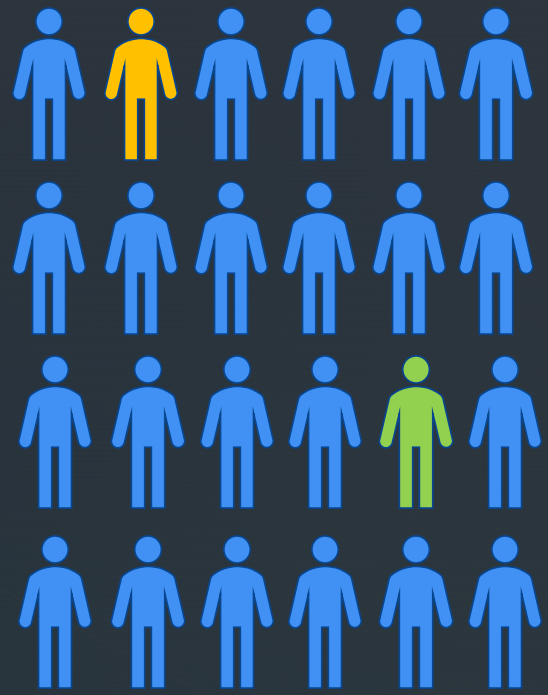
Birthday Problem

"Wie hoch ist die Wahrscheinlichkeit, dass von *n* zufällig ausgewählten Personen, mindestens zwei am selben Tag Geburtstag haben?"



- Konkretes Beispiel:

Wie hoch ist die Wahrscheinlichkeit,
dass bei 30 Personen,
mind. 2 am selben Tag Geburtstag
haben ?

Gliederung

- Geschichtlicher Hintergrund
- Annahmen für mathematische Modellierung
- Mathematische Herleitung
- Erweiterungen
- Warum Fehleinschätzung / Paradoxon?
- Anwendung Kryptographie
- Quellen

Geschichtlicher Hintergrund

Geburtstagsproblem wird 1927 Harald Davenport zugeschrieben, obwohl er es damals nicht publizierte

Kein Anspruch auf Entdeckung

"because he could not believe that it had not been stated earlier"

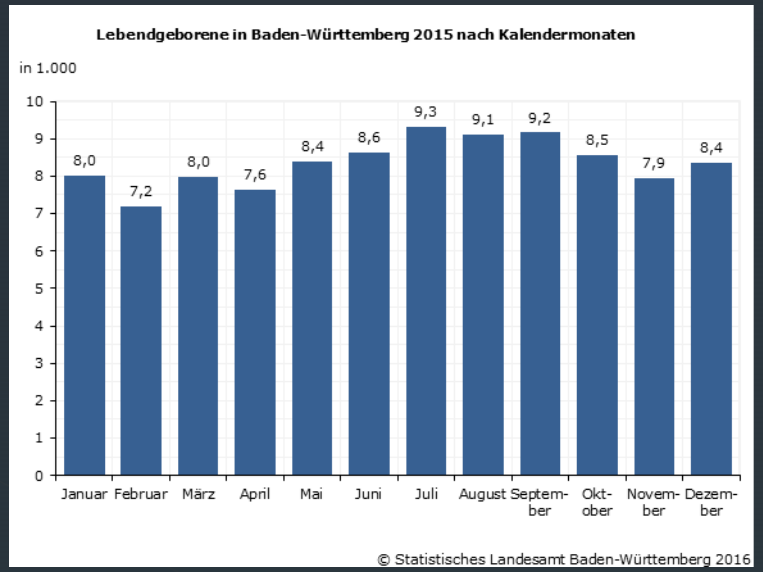
Problem erstmals publiziert von Richard von Mises im Jahr 1939 untersucht

Begriff "Geburtstagsparadoxon" wurde erst in den 1950er Jahren geprägt.

Heutzutage: Kryptographie, Wahrscheinlichkeitstheorie, Statistik und Informatik.

Annahmen für mathematische Modellierung

- Jahreszahl wird vernachlässigt
- Jedes Jahr hat 365 Tage
  Wahrscheinlichkeit wird geringfügig größer
- Alle 365 Tage eines Jahres sind als Geburtstage gleichwahrscheinlich.
  Wahrscheinlichkeit wird etwas kleiner.
- keine Zwillinge,...
- Die Auswahl der n Personen erfolgt hinsichtlich ihres Geburtstages „auf gut Glück“.



Mathematische Herleitung

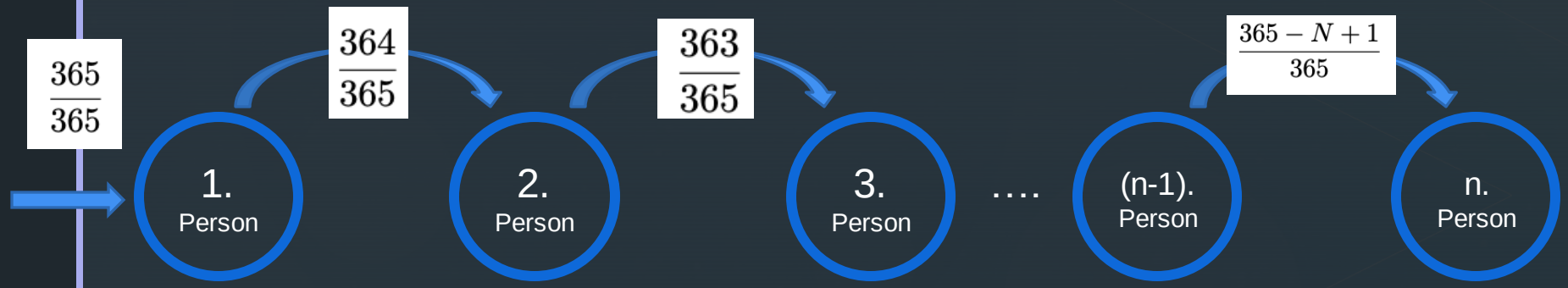
P(A) : Wahrscheinlichkeit, dass 2 Personen aus einer Gruppe am selben Tag geboren wurden.

➡ einfacher mit P(Ā) zu berechnen: Gegenwahrscheinlichkeit:

Alle Personen in dem Raum wurden an einem anderen Tag geboren.

Es gilt : P(A) = 1 – P(Ā)

Ā: nicht 2 oder mehr dürfen am selben Tag Geburtstag haben



mind. 2 Geburtstage am selben Tag:

$$= 1 - \frac{365 \cdot 364 \cdots (365 - N + 1)}{365^N}$$

$$= 1 - \prod_{i=1}^{N-1} \left(1 - \frac{i}{365}\right)$$

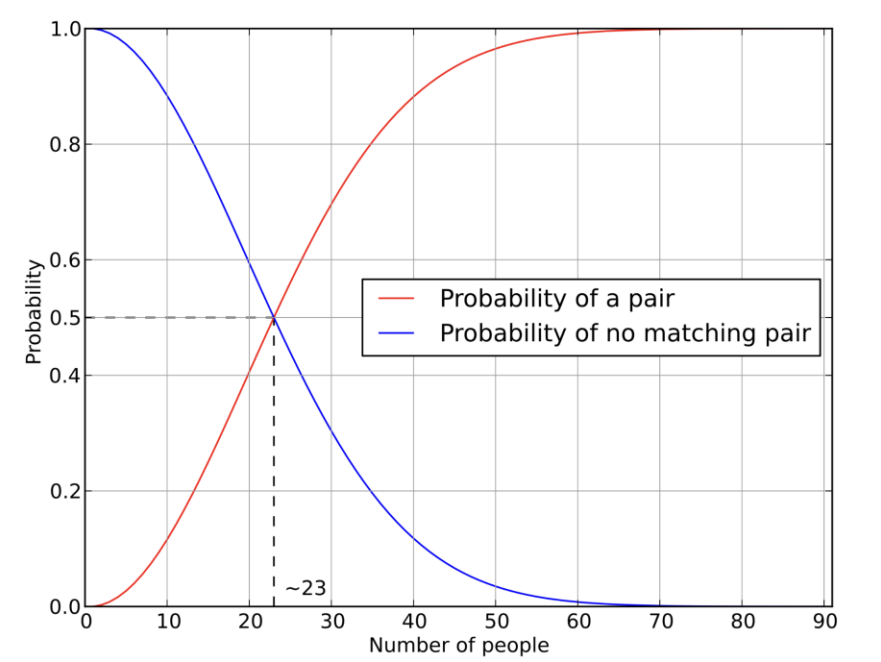
Allgemeine Formel:

c als Kategorie (hier Geburtstag)

$$1 - \prod_{i=1}^{N-1} \left(1 - \frac{i}{c}\right).$$

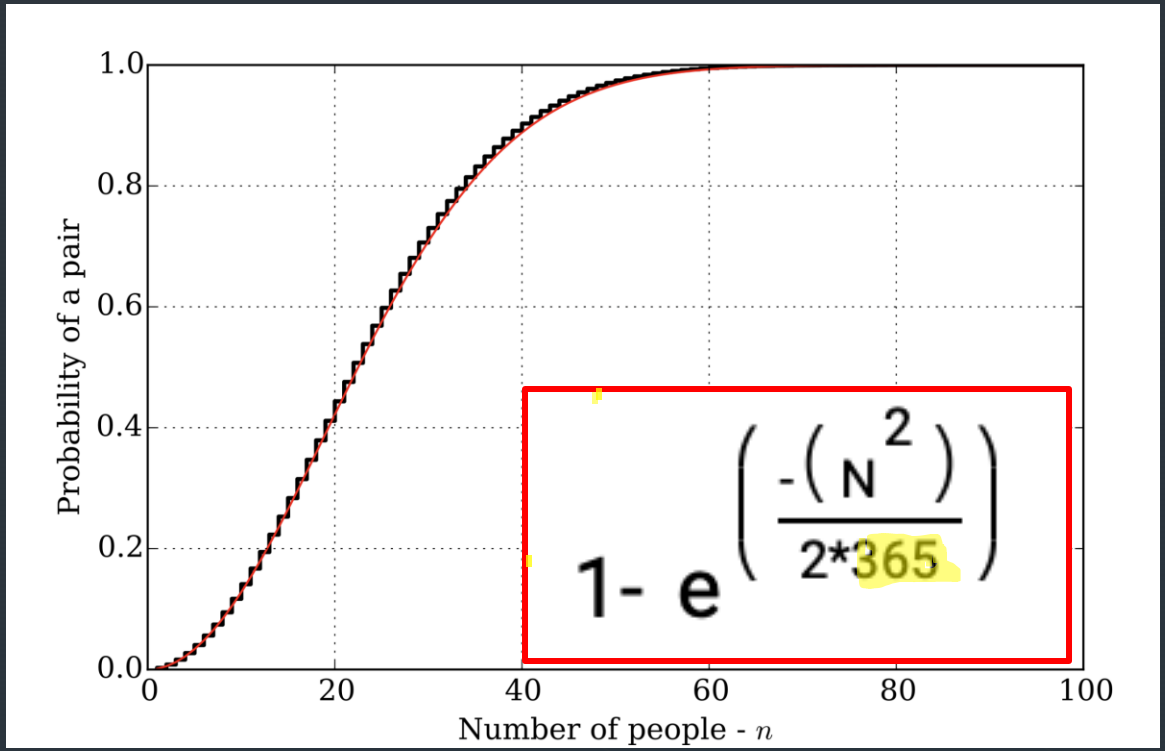
Auflösung

2 Leute aus einer Gruppe mit N Personen haben den gleichen Geburtstag:



n	$p(n)$
1	0.0%
5	2.7%
10	11.7%
20	41.1%
23	50.7%
30	70.6%
40	89.1%
50	97.0%
60	99.4%
70	99.9%
75	99.97%
100	99.999 97%
200	99.999 999 999 999 999 999 999 999 9998%
300	$(100 - 6 \times 10^{-80})\%$
350	$(100 - 3 \times 10^{-129})\%$
365	$(100 - 1.45 \times 10^{-155})\%$
≥ 366	100%

Annäherung



$$1 - e^{\left(\frac{-(N^2)}{2*c}\right)}$$

Annäherungsformel für N:

$$N \doteq 1.2\sqrt{c}$$

50 % Chance für 1 Treffer

Bsp.: $N = 1.2\sqrt{365} = 22.9260... \approx 23$

$$N \doteq 2.5\sqrt{c}$$

95 % Chance für 1 Treffer → 48

Erweiterung durch Kategorien

$$N \doteq 1.2 \sqrt{1 \Big/ \left(\frac{1}{c_1} + \frac{1}{c_2} + \cdots + \frac{1}{c_k} \right)}.$$

50 % Chance auf 1 Treffer

Anzahl innerhalb der unabhängigen Kategorien:

c_1 = Geburtstagstage = 365

c_2 = Wohnort = 1000

c_3 = Alter = 100

$$N \doteq 1.2 \sqrt{1 \Big/ \left(\frac{1}{365} + \frac{1}{1000} + \frac{1}{100} \right)} = 10.2375... \approx 10$$

Erweiterung durch k Treffer

$$Ne^{-N/ck}/(1 - N/c(k + 1))^{1/k}$$
$$= \left[c^{k-1}k! \log_e\left(\frac{1}{1-p}\right)\right]^{1/k}$$

~Paper: Methods for Studying Coincidences

Könnte
man numerisch lösen

recurrence relation

$$Q_k(N,c) = \sum_{i=1}^{\lfloor N/k \rfloor k} \left[\frac{N!c!}{c^{ik}i!(k!)^i(N-ik)!(c-i)!} \sum_{j=1}^{k-1} Q_j(N-ik, c-i)\frac{(c-i)^{N-ik}}{c^{N-ik}}\right]$$

3 Treffer:

$$P_3(N,c) = 1 - Q_1(N,c) - Q_2(N,c) = 1 - c^{-N}c!{}_2\tilde{F}_1\left(\frac{1}{2}N, \frac{1}{2}(1-N); 1+c-N; 2\right),$$

durch "regularized hypergeometric function" berechenbar:

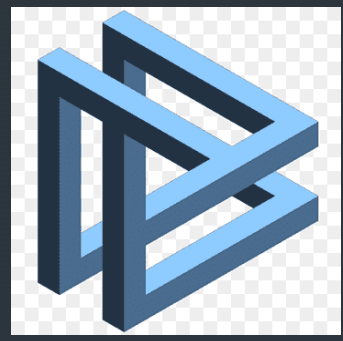
Kommt aus Statistik,
hypergeometrische Verteilung,
Funktion ist auch in R definiert

Mindestanzahl n von Personen	
k	n
2	23
3	88
4	187
5	313
6	460
7	623
8	798
9	985
10	1181

> 50 % Chance auf k Treffer, wie groß muss die Gruppe sein?

?

Warum Fehleinschätzung / Paradoxon?



- Interpretation/ Intuition: eine bestimmte Person hat an einem bestimmten Tag

- Intuition: Man geht durch den Rest der Gruppe und "prüft" auf Übereinstimmung

  pro Person Wkeit. ca. $\frac{1}{365}$

- Für Gruppe von 30 Personen: → $\frac{30}{365}$ oder etwa 8 % beträgt.

- Umfragen ergeben: meisten Menschen schätzen Intuitiv zwischen 1 und 10 %

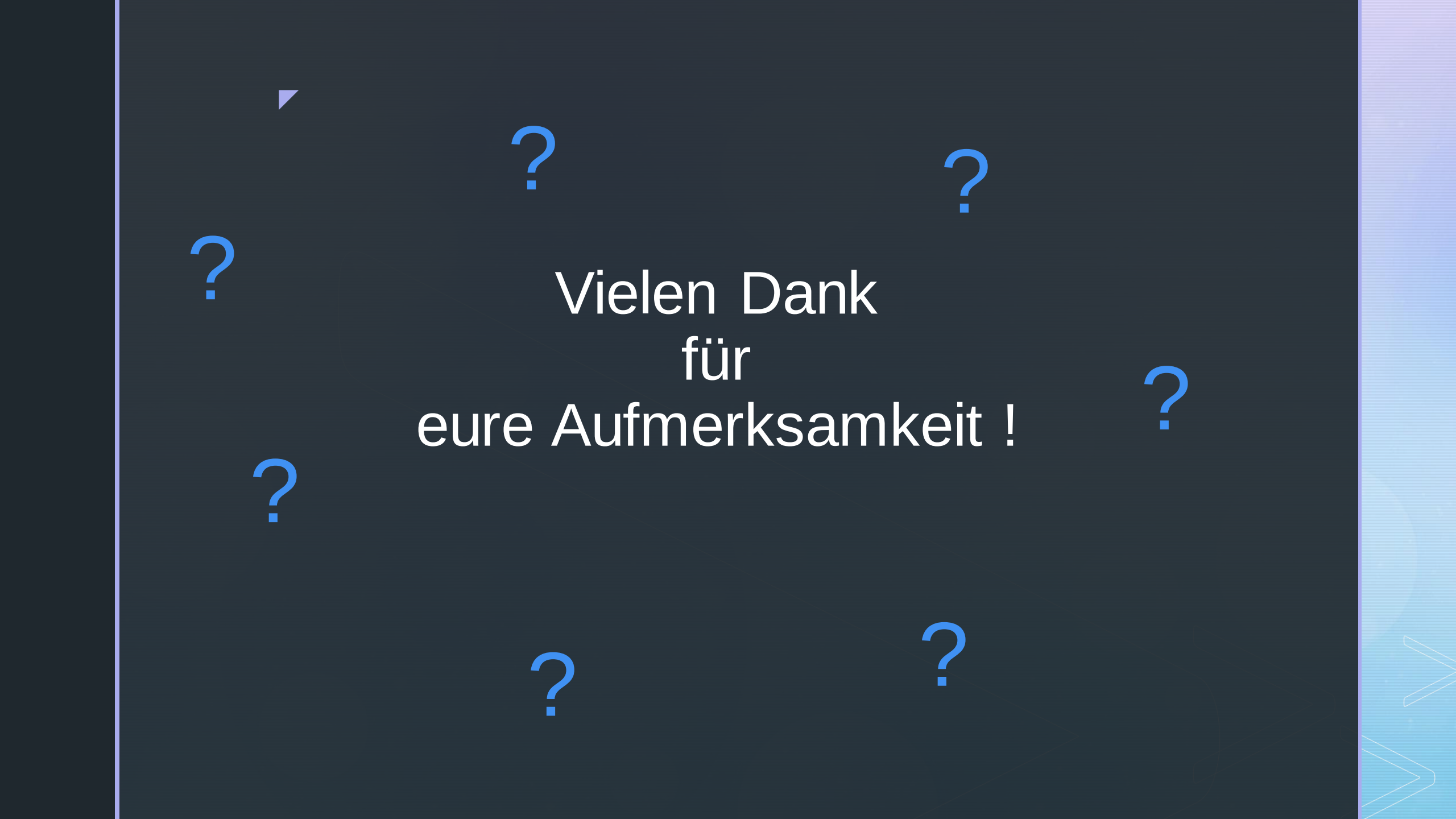
Das Paradoxon ist ein Beispiel dafür, wie die menschliche Intuition uns in die Irre führen kann, wenn es um die Schätzung von Wahrscheinlichkeiten geht

Anwendung Kryptographie

- Geburtstagsparadoxon: Grundlage für Angriffe Hashfunktionen verwendet.
  - mathematische Funktion, die eine beliebig große Menge von Daten auf eine feste Größe reduziert. (Passwörter)
  - berechnet Wkeit, dass 2 Eingabemengen gleiche Hashfunktion ergeben
  - gibt man ausreichend Anzahl an Eingabemengen in die Hashfunktion,

  ⟹ Wkeit., dass mind 2 Eingabemengen denselben Hashwert besitzen groß
- "Geburtstagsangriff"
- um Hashfunktionen zu brechen und die Sicherheit von Passwörtern und anderen kryptografischen Systemen zu gefährden.

Vielen Dank
für
eure Aufmerksamkeit !



Quellen:

- https://mathworld.wolfram.com/BirthdayProblem.html
- https://en.wikipedia.org/wiki/Birthday_problem
- https://joelvelasco.net/teaching/249/Diaconis%20and%20Mosteller%201989%20-%20methods%20for%20studying%20coincidences.pdf
- https://www.pngwing.com/de/free-png-sgbtt
- https://www.datenschutz-notizen.de/36c3-was-ist-eigentlich-kryptographie-0324431/
- https://de.wikipedia.org/wiki/Verallgemeinerte_hypergeometrische_Funktion
- https://www.mathematik.ch/anwendungenmath/wkeit/geburtstag/
- https://www.lernhelfer.de/schuelerlexikon/mathematik-abitur/artikel/das-geburtstagsproblem
- https://matheguru.com/stochastik/geburtstagsproblem.html