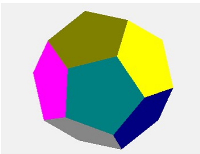
The birthday problem
And other
Improbable probabilities

This is an article from my home page: www.olewitthansen.dk



Ole Witt-Hansen 2009 (2016)

Contents

1. The Birthday Problem

It is well known that the theory of probabilities offers several examples of calculated probabilities, which at a glance appears to be quite improbable.
One of the most notorious is the birthday problem. It may for example be presented this way. Someone offers you a bet, whether or not two students in a class of 24 have birthday on the same day of the year.
Or, if we put it in more formal manner. If you pick 24 numbers at random from the numbers 1..365, is it then more likely that the same number will be drawn twice or more, than the numbers are all different?

To deal with this, we shall first determine the number of outcomes when drawing at random *n* times a number among 1…365. Since each drawing is independent of the others, the multiplication rule gives for the number of outcomes:

(1.1) $$N(U) = 365\cdot 365\cdot 365\cdot\ldots\cdot 365 = 365^n$$

Let H be the event: At least two persons among n person have birthday on the same day of the year. And let $\overline{H}$ be the complementary event: No one among the n persons have birthday on the same day of the year.

But instead of calculating the sought probability $P(H)$, we calculate the probability $P(\overline{H})$ that none of the n persons have birthday on the same day of the year. So we first calculate the total numbers of outcomes of that event:

If there are to be no coincidences, the possibility for choices of birthday for the first person is 365, for the second it is 364, and so on, so the total number of possibilities are.

$$N(\overline{H}) = 365\cdot 364\cdot 363\cdot\ldots\cdot(365-n+1) \qquad \text{(n - factors)}$$

The probability $P(\overline{H})$ is as usual calculated as the number of events belonging to $\overline{H}$, which is $n(\overline{H})$ divided by the total number of events $n(U)$

(1.2) $$P(\overline{H}) = \frac{n(\overline{H})}{n(U)} = \frac{365\cdot(365-1)\cdot(365-2)\cdot\ldots(365-n+1)}{365\cdot 365\cdot 365\cdot 365\cdot\ldots\cdot 365} = \frac{365}{365}\cdot\frac{365-1}{365}\cdot\frac{365-2}{365}\cdot\ldots\cdot\frac{365-n+1}{365}$$

Having some patience, one may calculate this probability with a pocket calculator for $n = 24$, and observe that it is a little less than 0.5, and consequently the somewhat surprising result is that the probability of a coincidence of birthdays among 24 persons exceeds 50%.

We shall now show that it is possible (applying a little mathematics) to give a simple expression for the probability for an arbitrary *n*.

To do so, we divide the denominator 365 up in the numerator in each of the fractions in (1.2), followed by taking the natural logarithm on each side.

(1.3) $$P(\overline{H}) = \frac{365}{365} \cdot \frac{365-1}{365} \cdot \frac{365-2}{365} \cdot ... \frac{365-n+1}{365} = 1 \cdot (1-\frac{1}{365}) \cdot (1-\frac{2}{365}) \cdot ... \cdot (1-\frac{n-1}{365})$$

(1.4) $$\ln P(\overline{H}) = \ln(1-\frac{1}{365}) + \ln(1-\frac{2}{365}) + ... + \ln(1-\frac{n-1}{365})$$

Expanding the natural logarithm to the first order, we have the formula:

$$\ln(1+h) \approx h \quad \textit{for} \quad h << 1 \quad \text{(meaning } h < 0.1\text{)} \;\; (\ln (1.1) = 0.095)$$

Applying this formula in each term in (1.2)

$$\ln P(\overline{H}) = -\frac{1}{365} - \frac{2}{365} - \frac{3}{365} - ... \frac{n-1}{365} = -\frac{1}{365}(1+2+3+...(n-1))$$

(1.5)

$$\ln P(\overline{H}) = -\frac{1}{365}\frac{n(n-1)}{2}$$

Our aim is then to determine n such that:

$$P(\overline{H}) < \tfrac{1}{2} \quad \Leftrightarrow \quad \ln P(\overline{H}) < \ln \tfrac{1}{2} \quad \Leftrightarrow \quad \ln P(\overline{H}) < -0.6931$$

which, gives the inequality:

$$-\frac{1}{365}\frac{n(n-1)}{2} < -0.6931 \quad \Leftrightarrow \quad n(n-1) > 505.96$$

This is a algebraic inequality of second order:

$$n^2 - n - 505{,}963 > 0$$

The solution to the corresponding quadratic equation is $n = \frac{1 \pm 45}{2} = \begin{pmatrix} 23 \\ -22 \end{matrix}$

And it gives solution to the inequality:

$$n^2 - n - 505{,}963 > 0 \quad \Leftrightarrow \quad n < -22 \quad \vee \quad n > 23$$

Everything so, when the number of persons is bigger than 23, the probability for no coincidence of birthday between them is less than 50%, meaning that the probability of a least one coincidence of birthday is greater than 50% when the number of persons are greater than 23.

2. The coin in one of the three boxes – problem

There exists several versions of this riddle, but a simple one goes as follows.

We have three identical boxes. One of them contains a (gold) coin.

A person is challenged to choose one. After having made the choice, one of the other two boxes that does not contain the coin, is removed.
The person is then asked, if he wants to alter his first choice among the two remaining boxes.

The question to decide is then: Is it advantageous to do a reselection or not, or is it insignificant?

The answer is, however, that he should always reselect, and the explanation is straightforward.

This is not really a combinatorial problem, but it is stated in a manner apt to confuse students. But that it also has caused lengthy discussions and minor articles among teachers in high school is incomprehensible.

The first choice of a box containing the coin is done with probability 1/3, which means that the probability that the box with the coin is among the other two boxes is 2/3.

These probabilities are unchangeable, if no other information is given.

But since only one of the two remaining boxes is left, the probability that it contains the coin is still 2/3.

One should therefore always make a new choice, since it gives twice the probability to get the box with the coin.

On the other hand, if one of the other two boxes were removed at random, there is no advantage of reselecting.

3. The number of permutations of n element with no fixed elements

In connection with the old children card game “war”, where two persons have a deck of cards, or just 13 cards in the same suit, I became aware of a similar problem as the birthday problem.

The two players each present one card from their decks. A “war” occurs when the two players lay down the same card.

The question then arises. What is the probability of a coincidence, when the two persons have 52 or 13 or n identical cards, and present one card at a time.

This is an example of a general problem of determining the number of permutation of n elements, where no element corresponds to itself. We shall denote this number of permutations $Q(n)$.
Since the number of permutation of n different elements is

(3.1) $$P(n) = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

Then the probability of coincidence is less than 50% if $Q(n) > 0.5 \cdot P(n)$, since the number of permutations, where no element corresponds to itself is larger than the total number of permutations.

Our aim is to determine $Q(n)$, the number of permutations with no fixed points.

However, this turns out to be much harder, than solving the birthday problem. Combinatorial problems may be very tricky, and in this sense, this problem is an excellent example.

As we are about to show, the probability of having a permutation with no fixed points converges to $1/e = 0.3679$ when the number n goes to infinity. However the limit is almost reached for $n = 5$. This means that the probability of a coincidence is $1 - 0.3679 = 0.6321$, a rather heavy overweight, when placing a bet.

3.1 making a recursion formula for *Q*(*n*)

It is actually quite unproblematic to make a recursion formula for $Q(n)$, where $Q(n)$ is expressed by all the previous $Q(n-1), Q(n-2), ..., Q(2)$

As usual we denote $P(n) = n!$ for the total number of permutations of n elements.

The number of permutations of n elements, where k elements are identical is:

(3.1) $$P_k(n) = \frac{n!}{k!},$$

since the permutations of the k identical elements do not change the permutation.

The number of q-combinations taking from a set of n different elements is:

(3.2) $$C(n,q) = \binom{n}{q} = \frac{n!}{(n-q)!q!}$$

$Q(n)$ denotes the number of permutations of n element, where no element is mapped into itself.

We establish the recursion formula in the following way:

From the number of all permutations $n!$, we subtract the permutations, where exactly one element corresponds to itself, the permutations, where exactly two element correspond to themselves,… the permutations, where exactly n-1 element correspond to themselves, the permutations, where all elements correspond to themselves.

If q elements correspond to themselves, they can be selected in $C(n,q)$ different ways. The remaining $n - q$ elements, can be permuted in $Q(n-q)$ different ways, such that no element corresponds to itself. We therefore have:

(3.3) $$Q(n) = n! - \binom{n}{1}Q(n-1) - \binom{n}{2}Q(n-2) - .... - \binom{n}{n-2}Q(2) - 1$$

The formula, cannot be applied for $n = 2$, but here we know, that $Q(2) = 1$, since there is only one permutation of two elements, where no element correspond to itself, namely the one where the two elements are switched. For $n = 3.. 5$, we find.

$Q(3) = 3! - 3\cdot Q(2) - 1 = 2$

$Q(4) = 4! - 4\cdot Q(3) - 6\cdot Q(2) - 1 = 9$

$Q(5) = 5! - 5\cdot Q(4) - 10\cdot Q(3) - 10\cdot Q(2) - 1 = 5! - 5\cdot 9 - 10\cdot 2 - 10\cdot 1 - 1 = 120 - 45 - 20 - 10 - 1 = 44.$

Although the recursion formula is applicable, it would be more satisfactory to have a formula for $Q(n)$, which do not imply calculating all the previous $Q(k)$, where $k = 1, 2,\ldots,n-1$, but as mentioned earlier it involves some cleaver tricks.

If we isolate $n!$ in the recursion formula (3.3), it can be written:

(3.4) $$n! = \sum_{k=0}^{n} \binom{n}{k} Q(n-k)$$

Dividing by $n!$

(3.5) $$1 = \frac{1}{n!}\sum_{k=0}^{n} \binom{n}{k} Q(n-k) = \sum_{k=0}^{n} \frac{Q(n-k)}{k!(n-k)!} = \sum_{k=0}^{n} \frac{Q(k)}{(n-k)!k!}$$

We shall then make use of this formula, the Taylor expansion of e^x , as well as for the formula for an infinite geometric series with quotient x.

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad \text{and} \qquad \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad \text{for } -1<x<1$$

We then look at the expression

(3.6) $$e^x \left(\sum_{q=0}^{\infty} \frac{Q(q)}{q!} x^q \right) = \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \right) \left(\sum_{q=0}^{\infty} \frac{Q(q)}{q!} x^q \right) = \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \right) R(x)$$

The first factor is e^x on both sides, and we have denoted the second factor as $R(x)$. When multiplying the two (infinite) sums, we collect all terms were $k + q = n$, so these terms all have the same factor $x^{k+q} = x^n$. Each term with exponent n can then be written.

(3.7) $$\sum_{k=0}^{n} \frac{Q(n-k)}{k!(n-k)!} x^n = x^n \sum_{k=0}^{n} \frac{Q(n-k)}{k!(n-k)!} = x^n$$

Where the last equation follows from (3.5): $\sum_{k=0}^{n} \frac{Q(k)}{(n-k)!k!} = 1$

Then taking the sum for $n = 0, 1, 2,\ldots\ldots$ The product of the two sums becomes:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

When this is inserted back in (3.6) we get:

$$e^x R(x) = e^x \sum_{q=0}^{\infty} \frac{Q(q)}{q!} x^q = \frac{1}{1-x} \quad \Leftrightarrow$$

(3.8) $$\sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^n = \frac{e^{-x}}{1-x}$$

Multiplying by 1- x and comparing terms with the same exponent of x gives:

$$\sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^n = \frac{e^{-x}}{1-x} \quad \Leftrightarrow$$

$$(1-x)\sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^n = e^{-x} \quad \Leftrightarrow \quad (1-x)\sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n$$

(3.9) $$\sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^n - \sum_{n=0}^{\infty} \frac{Q(n)}{n!} x^{n+1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n$$

Comparing terms with the same exponent n, we get:

(3.10) $$\frac{Q(n)}{n!} - \frac{Q(n-1)}{(n-1)!} = \frac{(-1)^n}{n!} \quad \Leftrightarrow \quad \frac{Q(n)}{n!} = \frac{Q(n-1)}{(n-1)!} + \frac{(-1)^n}{n!} \quad \Leftrightarrow$$
$$Q(n) = nQ(n-1) + (-1)^n$$

Since $Q(1)=0$, it follows that $Q(2)=1$, $Q(3)=2$; $Q(4)=9$, $Q(5)=44$, $Q(6)=265$;

Furthermore we get for the probability $P(n) = \frac{Q(n)}{n!}$ that we have a permutation where no element corresponds to itself.

$$\frac{Q(2)}{2!} = \frac{Q(1)}{1!} + \frac{(-1)^2}{2!} = \frac{(-1)^2}{2!}$$
$$\frac{Q(3)}{3!} = \frac{Q(2)}{2!} + \frac{(-1)^3}{3!} = \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!}$$
$$\frac{Q(4)}{4!} = \frac{Q(3)}{3!} + \frac{(-1)^4}{4!} = \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \frac{(-1)^4}{4!}$$

And in general:

(3.11) $$P(n) = \frac{Q(n)}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

From which it is seen that:

(3.12) $$P(n) \to \frac{1}{e} \ \ for \ \ n \to \infty$$

However, n does not need to be very large using this formula, since 1/e = 0.3679, which is compared to the values of $P(n)$ for n =1..5

$$P(2) = \frac{Q(2)}{2!} = \tfrac{2}{6} = 0.50 \ , \quad P(3) = \frac{Q(3)}{3!} = \tfrac{2}{6} = 0.33$$
$$P(4) = \frac{Q(4)}{4!} = \tfrac{3}{8} = 0.375 \ , \ P(5) = \frac{Q(5)}{5!} = \tfrac{44}{120} = 0.367$$

So already for n = 4, the result for $P(n)$ is close to 1/e = 0,368.

The recursion formula (3.10) $Q(n) = nQ(n-1) + (-1)^n$ for the number of permutations, where no element corresponds to itself, was first brought forward (by direct reason) by Montmort i 1708. Later it has been shown that his formula leads to series expansion above.
However, his argument is a little tricky.

We make a partition of the permutations of the numbers 1..n after the k'th element, where k = 2..n. Firstly we focus on the permutations without fixed points, where the element "1" is placed in the k'th place. The resting n-1 elements can be permuted on $Q(n$-1) different ways, having no fixed points.
Then we look at the permutations where "1" is not on the k'th place, but on the places 2..k - 1,k+1,..,n and k is on the first place. The remaining n – 1 elements can be permuted in $Q(n$-2) different ways without fixed points.
Since k takes the values 2..n, there are n – 1possibilities for k.

We can therefore conclude that we get all permutations without fixed points, the equation holds true.

(3.13) $$Q(n) = (n\text{-}1)(Q(n\text{-}1)+Q(n\text{-}2))$$

This recursion formula can, however, be simplified to:

(3.14) $$Q(n)=nQ(n\text{-}1)+(\text{-}1)^n$$

This we prove by induction.
The formula is correct for n = 2. We assume that the formula is correct for n, and prove that it also holds for n + 1. We insert

$$Q(n)=nQ(n\text{-}1)+(\text{-}1)^n \quad \Leftrightarrow \quad nQ(n\text{-}1)= Q(n) \text{-}(\text{-}1)^n$$

on the right side of the recursion formula (3.13) $Q(n) = (n$-1)($Q(n$-1)+$Q(n$-2))
using n +1 instead of n.

$$Q(n+1) = n(\mathrm{Q}(n)+ Q(n\text{-}1)) = nQ(n)+nQ(n\text{-}1)$$

(3.14) $= nQ(n) + Q(n) - (-1)^n = (n+1)Q(n) + (-1)^{n+1}$

Which proves the assertion.
If we divide the simplified formula (3.14) by $n!$, again we find the probability of having a permutation without fixed points

(3.15) $$P(n) = \frac{Q(n)}{n!} = \frac{Q(n-1)}{(n-1)!} + \frac{(-1)^n}{n!}$$

We see that the following equation is valid.

$$P(n) - P(n-1) = \frac{Q(n)}{n!} - \frac{Q(n-1)}{(n-1)!} = \frac{(-1)^n}{n!}$$

Evaluating the first terms.

$$P(1) = 0, \quad P(2) = \tfrac{1}{2!}, \quad P(3) = \tfrac{1}{2!} - \tfrac{1}{3!}, \; P(4) = \tfrac{1}{2!} - \tfrac{1}{3!} + \tfrac{1}{4!}$$
$$P(n) = \tfrac{1}{2!} - \tfrac{1}{3!} + \tfrac{1}{4!} + \ldots + \tfrac{(-1)^n}{n!}$$
$$P(\infty) = e^{-1}$$

We note that we get the same formula as (3.11)

4. The Sct. Petersborg paradox

The Sct. Petersborg paradox is first brought forward by Daniel Bernuilli in connection with theory of gambling.
This game is an example, where the theoretical calculated mean gain is infinitely high, but at the same time no sane person would ever dream of playing the game. The snag is of course that one condition for the calculated probabilities is the existence of infinite capitals.

Such paradoxes also appear in economic theory.

Even if an infinite capital does not exist, and is replaced by a cap on stakes, one may still draw similar conclusions, but they turn out to be considerably more discouraging for the player.

The game is indeed very simple. You play with a coin (or black/white on the roulette).
The player makes a bet with the bank, where the player deposits S units. The amount S is fixed by the bank. So the game is this: If the player hits N heads (or tails) in a row, he wins 2^N units.

The probability of scoring N heads followed by a tail is:

(4.1) $$P(N) = (\tfrac{1}{2})^N \, \tfrac{1}{2} = (\tfrac{1}{2})^{N+1}$$

The gain when throwing heads N times in a row is: $G(N) = 2^N - S$. The mean of the gain is therefore:

$$\text{(4.2)}\quad \begin{aligned} E(G) &= \sum_{N=0}^{\infty} P(N)G(N) = \sum_{N=0}^{\infty} (\tfrac{1}{2})^{N+1}(2^N - S) \\ &= \sum_{N=0}^{\infty} (\tfrac{1}{2} - (\tfrac{1}{2})^{N+1} S) = \infty - S\,\tfrac{1}{2}\frac{1}{1-\tfrac{1}{2}} = \infty - S = \infty \end{aligned}$$

It may come as a surprise that the mean gain actually is infinite, independently of the stake, and a Casino which offers many of these games, should certainly not risk to go bankrupt, unless the stake S is very, high in which case no one will play the game. That is the paradox.

The paradox does not come from probability though, but from the lack of symmetry between the Casino on one side, and say 10.000 players on the other
The Casino alone bears its gain or loss, but the loss among the players is distributed between e.g. 10.000, whereas the gain belongs to one only, unless 10.000 players agree to share the gain, but this hits a psychologically barrier, since no one would dream of sharing an extraordinary gain with 10.000 others.

4.1 The game with a cap over the gain[iii]

We shall now analyze the game, but this time where the bank has put a cap $M = 2^{N_0}$ over the gain, meaning that if the player hits heads N_0 times or more in a row, the gain is still M.

We then proceed to calculate the new mean of the gain, but with the changed rules.

$$\text{(4.3)}\quad E(G) = \sum_{N=0}^{\infty} P(N)G(N) = \sum_{N=0}^{N_0-1} P(N)G(N) + \sum_{N=N_0}^{\infty} P(N)(M - S)$$

$$E(G) = \sum_{N=0}^{N_0-1} (\tfrac{1}{2})^{N+1}(2^N - S) + (M - S)\sum_{N=N_0}^{\infty} (\tfrac{1}{2})^{N+1} = \tfrac{1}{2}N_0 - \tfrac{1}{2}S\frac{1-(\tfrac{1}{2})^{N_0}}{1-\tfrac{1}{2}} + (M - S)(\tfrac{1}{2})^{N_0+1}\frac{1}{1-\tfrac{1}{2}}$$

$$\text{(4.4)}\quad \begin{aligned} E(G) &= \tfrac{1}{2}N_0 - S(1-(\tfrac{1}{2})^{N_0}) + (M - S)(\tfrac{1}{2})^{N_0} \\ E(G) &= \tfrac{1}{2}N_0 - S + M(\tfrac{1}{2})^{N_0} \end{aligned}$$

Inserting $M = 2^{N_0}$, we find:

$$\text{(4.5)}\quad E(G) = \tfrac{1}{2}N_0 - S + 1\,.$$

The outcome is (remarkably) simple. If the Casino will insure a gain in the long run, they should just make certain that $E(G) < 0$, implicating that:

$$S > \tfrac{1}{2}N_0 + 1.$$

If the bank has a cap, corresponding to $N_0 = 16$, so that $M = 2^{16} = 65{,}536$ chips, then the stake should only be $S > 8 + 1$ chips, insuring that the game is in favour of the Casino.

This may seem a rather modest stake for the player, but his chance of winning becomes correspondingly smaller.

(4.6) $$P(M)=\sum_{N=N_0}^{\infty}(\tfrac{1}{2})^{N+1}=(\tfrac{1}{2})^{N_0+1}\frac{1}{1-\frac{1}{2}}=(\tfrac{1}{2})^{N_0}=\frac{1}{2^{16}}\quad for\ \ N_0=16$$

The condition that a player has a gain by throwing *N* heads in a row is therefore:

(4.7) $$\mathrm{S}<2^{\mathrm{N}}\ \Leftrightarrow\ N>\log_2 S$$

The probability for this event, is calculated above, where we shall just replace N_0 with *N*.

$$P(S<2^N)=(\tfrac{1}{2})^N=\tfrac{1}{S}\quad \text{(Where we have put } N = log_2S\text{)}$$

A remarkably simple result, however, not particular attractive for a player, if S > 10 chips.

If we want to calculate the probability for the player to obtain a gain, which double the stake, we must have.

$$2\mathrm{S}<2^{\mathrm{N}}\ \Leftrightarrow\ N-1>\log_2 S$$

Which, gives the probability

$$P(S<2^{N-1})=(\tfrac{1}{2})^{N-1}=\tfrac{1}{2S}\ ,$$

And so on.