



Why

Here's a nice (surprising) example of computing an event probability. Consider the following question: In a group of n people, what is the chance that there is a pair of people who were born on the same day of the same month.

Example

Suppose we model the sample space as the set of all lists of length n in the set $\{1, \dots, 365\}$ —ignoring leap years. As usual, $|\Omega| = 365^n$. Suppose further that we model all sequences as equally likely. In other words, we define a distribution $p : \Omega \rightarrow [0, 1]$ by

$$p(\omega) = \frac{1}{365^n}$$

This is just one model of reality—some people may object to it since births may be less likely on certain days of the year, such as holidays.

It is easier to think of the event that all birthdays are distinct. Then, using the properties of event probabilities, we will be able to calculate the probability of the complement of this event. Define D_n so that

$$D_n = \{\omega \in \Omega \mid \omega_1 \neq \omega_2 \neq \dots \neq \omega_n\}$$

By the fundamental principle of counting, $|S_n| = 365 \cdot 364 \cdots (365 - n + 1)$. Thus the probability of the event D_n is

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

And so the probability of the event

$$S_n = \Omega - D_n = \{\omega \in \Omega \mid \exists i, j \text{ with } \omega_i = \omega_j\}$$

is

$$P(S_n) = 1 - P(D_n) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}$$

Even with small values of n , the above probability is quite large. For example, with n at 23, $P(S_{23}) \approx .51$. For example, with $n = 40$, $P(S_{40}) \approx .89$. Many people refer to the result of this particular probabilistic model as the *birthday paradox*. The word paradox is used to indicate that the probability is *higher* than one might expect. The reasoning for this is that very few people know someone with their own same birthday, even though we know many (> 40) people.

The resolution of the confusion is that if we fix the day of the year, we can consider the probability that in a group of n people there is an individual with the same birthday. The probability of the appropriate event under the same distribution as before can be shown to be¹

$$1 - \frac{364^n}{365^n}$$

Now, for $n = 40$, this value does not exceed 0.11.

¹Future editions will show.

