The Birthday Paradox

Andreas Klappenecker

Suppose that a group of people meet in a room. What is the probability that two of them have the same birthday?

We analyze a more general setting that has wider applicability. Suppose that there are n possible birthdays and k people in a room. We assume that the birthdays are uniformly and independently distributed over $N = \{1, \ldots, n\}$. If person a has its birthday at day b_a in N, then the event $(b_1, \ldots, b_k)$ has probability $1/n^k$.

We want to calculate the probability p of the event that two people have the same birthday. The complementary event is that all k people have different birthday, and we denote the probability of this event by $q = 1 - p$.

We calculate the probability q. Suppose that E is the subset of vectors in N^k that have pairwise distinct entries. Then

$$|E| = \prod_{i=0}^{k-1} (n-i),$$

because we have n possibilities to choose the first entry, $n - 1$ for the next entry, and so on. It follows that q is given by

$$q = \frac{|E|}{n^k} = \frac{1}{n^k} \prod_{i=0}^{k-1} (n-i) = \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right).$$

Recall that $1 + x \leq e^x$ holds for all real numbers x; hence,

$$q \leq \prod_{i=1}^{k-1} e^{-i/n} = \exp\left(-\sum_{i=1}^{k-1} \frac{i}{n}\right) = \exp\left(-\frac{k(k-1)}{2n}\right).$$

If

$$k \geq \frac{1}{2}(1 + \sqrt{1 + 8n \log 2}),$$

then $k(k-1) = \frac{1}{4}(8n \log 2) = n 2 \log 2$, thus

$$q \leq \exp(-k(k-1)/2) \leq 1/2.$$

Therefore, the probability $p = 1 - q$ that two people have the same birthday is at least $1/2$ when $k \geq \frac{1}{2}(1 + \sqrt{1 + 8n \log 2})$.

We followed [J.A. Buchmann, *Introduction to Cryptography*, Springer, 2001] in this presentation of the birthday paradox. An interesting application is the birthday attack on cryptographic hash functions. Our motivation of Pollard's ρ algorithm was also based on the birthday paradox.