The Birthday Paradox

September 18, 2003

1 Two Birthdays the Same

It turns out that there is at least a 50% chance that in any random sample of 23 people, two of them will have the same birthday. By "birthday", I mean you don't include the year; so, an example of a birthday would be "June 11".

The reason that one uses the "paradox" to refer to this phenomenon is that it seems counterintuitive that a random sample of so few people should likely have a matching pair of birthdays. The reason that such a small number of people suffices is that there are many pairs of individuals, and so many chances for a collision of birthdays: Indeed, with 23 people, there are $\binom{23}{2} = 253$ pairs of individuals.

Let us now prove that the probability is indeed 50%. First, let the sample space S be the set of all sequences $(x_1, ..., x_{23})$, where $1 \leq x_i \leq 365$. The value of x_i indicates the day of the year of the ith person's birthday. We are assuming here, for simplicity, that there are no leap years, so that each year has exactly 365 days. Another assumption we will make is that each day of the year is equally likely to be the birthday of a randomly selected person. It turns out that if we *don't* assume this, then the probability of a collision is even *greater* (with 23 people). So, with this assumption, we will get that any $s \in S$ has probability $P(s) = (365)^{-23}$. Note here that we are implicitly assuming that $\Sigma = 2^S$.

Now, let $E_2, ..., E_{23}$ be events defined as follows: Event E_i is the subset of S where person i's birthday is different from person j's birthday for all $j = 1, 2, ..., i-1$. Then, the event in S where all 23 people have *different* birthdays is easily seen to be $E_2 \cap E_3 \cap \cdots \cap E_{23}$, and so we want to calculate $P = P(E_2 \cap E_3 \cap \cdots \cap E_{23})$. The probability that two persons will have the same birthday will then be $1 - P$.

So, we just need to show that $P < 0.5$. If we knew that the E_i's were independent, then we could compute $P = P(E_2) \cdots P(E_{23})$, and then computing the $P(E_j)$'s separately would give us P. However, these events are not independent. Nonetheless, we may use the "chain rule" or "product rule" (not to be confused with the chain rule or product rule from calculus) for calculating the probability of events. This rule is given as follows:

Product Rule for Events.

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) \ = \ P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

In our case, we let $A_1 = E_2, ..., A_{22} = E_{23}$; and, for $j = 2, ..., 23$, we have that $P(E_j | E_1 \cap \cdots \cap E_{j-1})$ is the probability that person j's birthday is distinct from person 1's, person 2's,..., and person $(j-1)$'s, *given* that the birthdays of persons 1 through $(j-1)$ are all distinct. This probability is clearly $(366 - j)/365$. So, from the product rule we get

$$P(E_2 \cap \cdots \cap E_{23}) \ = \ \prod_{j=2}^{23} \frac{366 - j}{365} \ = \ \prod_{j=1}^{22} \left(1 - \frac{j}{365}\right) \ = \ 0.4927...$$

This is the probability that all the 23 birthdays are distinct; and so, the probability that two people *have* a common birthday is $1 - 0.4927...$, which exceeds 50%.