Math 302.102 Fall 2010
The Birthday Problem

Example. Suppose that there are n students in a class. What is the probability that at least two of those students have the same birthday?

In order to check whether or not this event has occurred in a particular class, we need some way of ordering the students. We can do this either by seating arrangement or alphabetically. Then, we can go through the list of students checking whether or not each birthday is one that has appeared previously. If a repeat birthday is found, then stop—there are at least two students in the class with the same birthday.

Let D_j be the event that the first j birthdays are different.

Let B_n be the event that at least two of the n students in the class have the same birthday.

We want to compute $\mathbf{P}\{B_n\}$. Notice that $B_n = D_n^c$ so that

$$\mathbf{P}\{B_n\} = 1 - \mathbf{P}\{D_n\}.$$

Here is the key assumption we'll make. *No matter what the birthdays of the first $j-1$ students, the birthday of the jth student is equally likely to be any one of the 365 days of the year.* (Although this ignores leap years and seasonal variations in the birth rates, etc., it can be shown that these considerations do not affect the answer very much.)

Notice that

$$\mathbf{P}\{D_2\} = \frac{364}{365} = \left(1 - \frac{1}{365}\right).$$

No matter what the birthday of the first student, there are 364 out of 365 possible birthdays for the second student which would make the first and second students have different birthdays. Furthermore,

$$\mathbf{P}\{D_{j+1} \,|\, D_j\} = \frac{365-j}{365} = \left(1 - \frac{j}{365}\right)$$

because, if we are given D_j, meaning the first j students have j different birthdays, then no matter what those birthdays are, the next student must have one of the remaining $365 - j$ birthdays. Also notice that if $i < j$, then $D_j \subset D_i$. This says that if the first j birthdays are different and i is less than j, then the first i birthdays must certainly be different. Hence,

$$D_j = D_1 \cap D_2 \cap \cdots \cap D_j.$$

By using the General Multiplication Rule, we find

$$\mathbf{P}\{D_n\} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right).$$

(See *Summary of Lectures from September 13, 2010, through September 27, 2010* handout.)

How can we estimate this probability?

In order to approximate this probability, we can take (natural) logarithms and use the fact that

$$\log(1+x) \approx x$$

for x near 0. (This is the first order Taylor's series approximation which is also called the tangent line approximation or linear approximation.)

That is,

$$\begin{aligned}
\log \mathbf{P}\{D_n\} &= \log\left(1-\frac{1}{365}\right) + \log\left(1-\frac{2}{365}\right) + \cdots + \log\left(1-\frac{n-1}{365}\right) \\
&\approx -\frac{1}{365} - \frac{2}{365} - \cdots - \frac{(n-1)}{365} \\
&= -\frac{1}{365}(1+2+3+\cdots+(n-1)) \\
&= -\frac{1}{365}\cdot\frac{(n-1)n}{2} \\
&= -\frac{n(n-1)}{2(365)}
\end{aligned}$$

implying that

$$\mathbf{P}\{D_n\} \approx \exp\left\{-\frac{n(n-1)}{2(365)}\right\} \quad \text{and} \quad \mathbf{P}\{B_n\} \approx 1 - \exp\left\{-\frac{n(n-1)}{2(365)}\right\}.$$

For example, if $n = 77$, then the probability that at least two students in the class have the same birthday is

$$\mathbf{P}\{B_{77}\} \approx 1 - \exp\left\{-\frac{77(76)}{2(365)}\right\} \doteq 0.99967.$$

Suppose that we want to have at least a 50% chance of finding a birthday match. How many students need to be in the class? We must choose n so that

$$\mathbf{P}\{B_n\} \geq 0.50 \quad \text{or} \quad 1 - \exp\left\{-\frac{n(n-1)}{2(365)}\right\} \geq 0.50.$$

Simplifying implies

$$\log 2 \leq \frac{n(n-1)}{2(365)} \leq \frac{n^2}{2(365)},$$

and so we need n to satisfy

$$n^2 \geq 2(365)\log(2) \quad \text{or} \quad n \geq \sqrt{2(365)\log(2)} \doteq 22.49.$$

Thus, a class with at least $n = 23$ students will have at least a 50% chance of producing a birthday match.