01204211 Discrete Mathematics
Lecture 6c: The pigeonhole principle and the birthday problem

Jittat Fakcharoenphol

August 23, 2022

The sock problem

I have n pairs of socks. Each pair is different from the other pair. How many socks do I have to pick out to be sure that I have at least one matching pair.

The Pigeonhole Principle

The answer of the previous question seems obvious. But it appears to be very useful in numerous cases. It is called **the pigeonhole principle**.

The pigeonhole principle

If we put $n + 1$ objects into n boxes, at least one box gets more than one objects.

Example

Assume that nobody is taller than 250 cm. In a group of 251 people, there are at least two people whose heights differ by at most 1cm.

Students with the same birthday[1]

- It is quite often that you find people with the same birthday.
- Since there is at most 366 days in a year, the pigeonhole principle states that if you have 367 people is a room, there is at least one pair with the same birthday.
- But that's the worst case scenario, as it is more common to find people with the same birthday. (In the next class, we will try to see if there is a pair of students in the class with the same birthday.)
- So, let's think about the probability that there are two students with the same birthday in a room with 40 students.

[1] This section follows section 2.5 in [LPV].

A simple case

- Let's start with 2 people in the room.
- **Notes:** While we have not defined properly what probabilities mean, we can count the number of all possibilities and the number of cases that we are interested in, and then calculate probability as the ratio between the two. E.g., if there are 50 possible outcomes and 30 of them are the ones we are interested in, the probability is 0.6. (Note that we assume that every outcome is equally likely.)
- How many possible birthdays can two people have?
  - Since each person has 366 choices, and the first person and the second person can choose independently, the number is $366 \cdot 366$.
- How many possible ways can they share the same birthday?
  - Since the first person has 366 choices, and the second person has to choose the same day, there are only 366 ways.
- Thus, the probability is $\frac{366}{366^2} = 0.0027$, very unlikely.

3 people

- Let's consider 3 people.
- How many possible birthdays can 3 people have?
  - Since each person has 366 choices, and each person can choose independently, the number is $366 \cdot 366 \cdot 366 = 366^3$.
- How many possible ways can at least two of them share the same birthday?
  - There are many cases.
  - So let's think about the case when everyone do not share any birthdays.
  - The first person has 366 choices. The second one has $366 - 1 = 365$ choices. The third one has $366 - 2 = 364$ choices. Thus, the number of ways they do not share any birthdays is $366 \cdot 365 \cdot 364$.
  - Notice that this is the number of ordered subsets.
- Thus, the probability that they do not share birthdays is $\frac{366 \cdot 365 \cdot 364}{366^3} = 0.9918$. Thus the probability that two of them share a birthday is $1 - 0.9918 = 0.0082$.

40 people

- Let's extend our previous argument to the case with 40 people.
- How many possible birthdays can 40 people have?
  - 366^{40}.
- How many possible ways that they do not sure any birthdays?
  - This is the number of ordered subsets with 40 elements of a 366-set.
  - Thus, there are $366 \cdot 365 \cdot 364 \cdots 327$ ways.
- Thus, the probability that they do not share birthdays is

$$\frac{366 \cdot 365 \cdot 364 \cdots 327}{366^{40}}.$$

- Umm... how small is it?
- Again you can use a computer to compute the exact value of this quantity. For example, you may want to use Wolfram Alpha.
- Anyway, we will try to estimate it using basic mathematical tools.

General case: n days k people

- Let's continue on the general case. When we have k people and a year contains n days, the probability that no two people share the same birthday is

$$\frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{n^k}.$$

- If this number is very close to 0, then it is very unlikely that no two people share the same birthday, i.e., it is very likely that there exists two people with the same birthday.

A few tweaks

- Dealing with small numbers is sometimes troublesome. (The reason will be more apparent later when we start introducing the tools.) So let's consider the reciprocal instead:

$$\frac{n^k}{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}.$$

- The top term looks easy to deal with; the bottom one does not. Let's break up the product:

$$\left(\frac{n}{n}\right) \cdot \left(\frac{n}{n-1}\right) \cdot \left(\frac{n}{n-2}\right) \cdots \left(\frac{n}{n-k+1}\right).$$

- If you look closely at this product, you can see that each term is at least one. In the beginning, the terms are very close to one and they get larger at the end.

The logarithms

- There is a nice tool that you can turn multiplications to additions: **logarithms.** So let's try to take the logarithms; we get

$$\ln\left(\left(\frac{n}{n}\right)\cdot\left(\frac{n}{n-1}\right)\cdot\left(\frac{n}{n-2}\right)\cdots\left(\frac{n}{n-k+1}\right)\right)$$

$$=\ln\left(\frac{n}{n}\right)+\ln\left(\frac{n}{n-1}\right)+\ln\left(\frac{n}{n-2}\right)+\cdots+\ln\left(\frac{n}{n-k+1}\right).$$

- The terms do not look that much better. But there's a nice fact about the natural logarithms.

$\ln x$: the upper bound

Fact:

$$\ln x \leq x - 1$$

This fact can be proved with elementary calculus. But it is fairly clear if you plot the functions $\ln x$ and $x - 1$.

$\ln x$: the lower bound

We know that

$$\ln x \leq x - 1$$

If we use the fact that $\ln \frac{1}{x} = -\ln x$, we can obtain the lower bound.

$$\ln x = -\ln \frac{1}{x} \geq -\left(\frac{1}{x} - 1\right) = \frac{x-1}{x}.$$

Let's conclude by stating the lemma:

Lemma 1

$$\frac{x-1}{x} \leq \ln x \leq x - 1.$$

The lower bound

Let's look at each term in the sum: $\ln\left(\frac{n}{n-j}\right)$. Using the lower bound in Lemma 1, we get that

$$\ln\left(\frac{n}{n-j}\right) \geq \frac{\frac{n}{n-j}-1}{\frac{n}{n-j}} = \frac{\frac{n-n+j}{n-j}}{\frac{n}{n-j}} = \frac{j}{n}.$$

Thus,

$$\ln\left(\left(\frac{n}{n}\right)\cdot\left(\frac{n}{n-1}\right)\cdot\left(\frac{n}{n-2}\right)\cdots\left(\frac{n}{n-k+1}\right)\right)$$

$$\begin{aligned}
&= \ln\left(\frac{n}{n}\right)+\ln\left(\frac{n}{n-1}\right)+\ln\left(\frac{n}{n-2}\right)+\cdots+\ln\left(\frac{n}{n-k+1}\right)\\
&\geq \frac{0}{n}+\frac{1}{n}+\frac{2}{n}+\cdots+\frac{k-1}{n}\\
&= \frac{1}{n}\left(1+2+\cdots+(k-1)\right) = \frac{k(k-1)}{2n}.
\end{aligned}$$

The upper bound

Again, let's look at each term in the sum: $\ln\left(\frac{n}{n-j}\right)$. Using the upper bound in Lemma 1, we get that

$$\ln\left(\frac{n}{n-j}\right) \leq \frac{n}{n-j} - 1 = \frac{j}{n-j}.$$

Thus,

$$\ln\left(\left(\frac{n}{n}\right) \cdot \left(\frac{n}{n-1}\right) \cdot \left(\frac{n}{n-2}\right) \cdots \left(\frac{n}{n-k+1}\right)\right)$$

$$\begin{aligned}
&\leq \frac{0}{n-0} + \frac{1}{n-1} + \frac{2}{n-2} + \cdots + \frac{k-1}{n-k+1} \\
&\leq \frac{0}{n-k+1} + \frac{1}{n-k+1} + \frac{2}{n-k+1} + \cdots + \frac{k-1}{n-k+1} \\
&= \frac{1}{n-k+1}\left(1+2+\cdots+(k-1)\right) = \frac{k(k-1)}{2(n-k+1)}.
\end{aligned}$$

Both

Using the derived upper and lower bounds, we get

$$e^{\frac{k(k-1)}{2n}} \leq \frac{n^k}{n(n-1)(n-2)\cdots(n-k+1)} \leq e^{\frac{k(k-1)}{2(n-k+1)}}$$

Let's plug in $n = 366$ and $k = 40$:

$$8.42 \leq \frac{366^{40}}{366 \cdot 365 \cdots 327} \leq 10.86.$$

So the probability that we get no two people with the same birthday is between $1/8.42 \approx 0.118$ and $1/10.86 \approx 0.092$. So we have high chance of finding two students with the same birthday. This is pretty close as the actual value is 0.1094.