2. Estimates: O, o-notation, Stirling formula, Birthday paradox and the bell curve

To read:
[1] 2.2.4. Pigeonhole principle. 2.2.5 The Twin Paradox
[3] 3.4. Estimates: an introduction - starting from 3.4.2. - Big Oh, little oh, 3.5.5. Estimate $n!$ - second proof only, 3.7. Inclusion - Exclusion.

2.1. O, o-notation.

Definition 2.1. Let $f, g : \mathbb{Z}_{\geq 0} \to \mathbb{R}$. We say that f is *big-Oh* of g and we write $f(x) = O(g(x))$ if there exist n_0 and c constants such that for all $n > n_0$, we have $|f(n)| < c \cdot |g(n)|$.

Definition 2.2. Let $f, g : \mathbb{Z}_{\geq 0} \to \mathbb{R}$. We say that f is *little-oh* of g and we write $f(x) = o(g(x))$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Examples: $n = O(n^2)$ and also $n = o(n^2)$, $n = O(2^n)$, $n = o(2^n)$, $\sin(n) = O(1)$ and $\sin(n)$ is not $o(1)$.

2.2. Stirling's formula.

Theorem 2.3. *(Stirling's formula)*

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

where $\sim$ is used to indicate that the ratio of the two sides tends to 1 as n goes to ∞.

2.3. Twin paradox.

Suppose that there are 50 students in a math class. What are the chances that two of them share the same birthday?

Theorem 2.4. *Suppose that $k \leq n$ are positive integers and each of k different people chooses 1 element from the set $[n]$. Their choices are uniformly random and independent. Then the probability $P = \frac{n!}{(n-k)!n^k}$ that they have chosen k different elements can be estimated as*

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

Proof. We will use the following inequality for $\ln(x)$.

Lemma 2.5. *For $x > 0$,*

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

Now we estimate

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

Also we find

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

$$\leq \left(\frac{n}{n-1} - 1\right) + \left(\frac{n}{n-2} - 1\right) + \ldots + \left(\frac{n}{n-k+1} - 1\right) = \frac{1}{n-1} + \frac{2}{n-2} + \ldots + \frac{k-1}{n-k+1}$$

$$\leq \frac{1}{n-k+1} + \frac{2}{n-k+1} + \ldots + \frac{k-1}{n-k+1} = \frac{1}{n-k+1}(1+2+\ldots+(k-1))$$

$$= \frac{k(k-1)}{2(n-k+1)}.$$

Applying the exponential function to both sides of our estimates we get the following:

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

□

So the answer to the question in the beginning of this paragraph is between 96.51% and 97.93%. More precisely, the probability is about 97.03%.

Now we will estimate the binomial coefficients. The binomial coefficients in the n-th row of the Pascal's triangle satisfy the following inequalities:

$$\binom{n}{0} < \binom{n}{1} < \cdots < \binom{n}{[n/2]}$$

and

$$\binom{n}{[n/2]+1} > \binom{n}{[n/2]+2} > \cdots > \binom{n}{1} > \binom{n}{0}.$$

Therefore, the middle binomial coefficient $\binom{n}{[n/2]}$ is the largest in the respective row. Stirling's formula implies that the largest binomial coefficient satisfies

$$\binom{n}{n/2} \sim \sqrt{\frac{2}{\pi n}}2^n.$$

Also we have the following formula describes how binomial coefficients decrease as we move away from the middle of the Pascal's triangle.

Proposition 2.6. *Let m, t be positive integers and $t \leq m$. Then*

$$e^{-t^2/(m-t+1)} \leq \frac{\binom{2m}{m-t}}{\binom{2m}{m}} \leq e^{-t^2/(m+t)}.$$

Proof. Here we prove the lower bound. We have

$$\frac{\binom{2m}{m}}{\binom{2m}{m-t}} = \frac{(m+t)(m+t-1)\cdots(m+1)}{m(m-1)\cdots(m-t+1)}.$$

It will be convenient for us to estimate the logarithm of this quantity.

$$\ln\left(\frac{(m+t)(m+t-1)\cdots(m+1)}{m(m-1)\cdots(m-t+1)}\right) = \ln\left(\frac{m+t}{m}\right) + \ln\left(\frac{m+t-1}{m-1}\right) + \ldots + \ln\left(\frac{m+1}{m-t+1}\right)$$

$$\leq \left(\frac{m+t}{m} - 1\right) + \left(\frac{m+t-1}{m-1} - 1\right) + \ldots + \left(\frac{m+1}{m-t+1} - 1\right) = \frac{t}{m} + \frac{t}{m-1} + \ldots + \frac{t}{m-t+1}$$

$$\leq \frac{t}{m-t+1} + \frac{t}{m-t+1} + \ldots + \frac{t}{m-t+1} = \frac{t^2}{m-t+1}.$$

Acknowledgements: I thank Prof. Janos Pach for designing this course and Dr. Matthew de Courcy-Ireland for sharing his lecture notes.

References

[1] Discrete Mathematics (L. Lovasz, J. Pelikan , K. Vesztergombi);
[2] Combinatorics: Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability (B. Bollobas);
[3] Invitation to Discrete Mathematics (J. Matousek, J. Nesetril).
[4] Extremal combinatorics (S. Jukna).
[5] Combinatorial theory (M. Hall), Blaisdell publishing company, 1967.