Ondřej Sladký

14. března 2022

Please do not share this task with anyone outside the seminar.

We are interested in the eye color of an unborn child. For simplicity we may assume that there are only two alleles for eye color – an allele for brown color and an allele for blue color and that they are both present in the population with the same probability.

A person has two allels for a gene for eye color (from biological point of view this is huge oversimplifcation). If at least one of them is brown, the person has brown eye color. Otherwise they have blue eyes.

We're given the child's family tree as a graph. The child is connected to its parants, they're connected to the respective grandpartents and so on.

Every person in the family tree has known eye color – either blue, brown or the color can be unknown. It may happen that at least one of the parents is missing. In such case, we assume that each of the alleles of the missing parent has the same probability to be brown as to be blue (i.e. that all the four possibilities have the same probability).

Note that even if we know the person's eye color we cannot completely determine their alleles. A person with brown eyes can have either both alleles brown or one of them blue and the other brown. If we know nothing about their predecessors, all of these three possibilities are equally likely and therefore this person has exactly $\frac{1}{3}$ probability to be a homozygote (i.e. to have both alleles brown).

In general, if we want to estimate the probability distribution of person's alleles, we first consider the probability distributions of their parents' alleles. The first allele of the descendant is taken randomly from the two alleles of their father. The second is taken randomly from their mother. With this process we obtain the prior probability distribution and we must update it to the knowledge of the descendat's eye color by considering only the cases which would result in the given eye color.

With this knowledge we should be able to estimate the probability that the unborn child has brown eyes.

However the format of the data representation described above can be quite ineffective, especially when there are a lot of family members whose eye color is unknown. That's why the given family tree is comprimed. A comprimed family tree has the same format as the uncomprimed one, but the edges are weighted. The weight of the edge corresmonds to the number of family members we know nothing about between the two nodes.

For example if we do not know the eye color of the child and neither the color of its father, but we know the eye color of the grandfather (it's brown), in the uncomprimed version we would have three verticies (one of them with brown eye color and two of them with unknown). In the comprimed version we have only two verticies connected with an edge of weight 1 (corresponding to the unknown father's eye color). In both representations, the probability that the child has brown eyes should be $\frac{19}{24}$.

The reason for this result is following: The grandfather has probability $\frac{1}{3}$ to be brown homozygote and probability $\frac{2}{3}$ to be heterozygote (to have one allele brown and the other blue) as explained above. The father's first allele has therefore $\frac{1}{3}$ probability to be blue (the rest to be brown) and his second allele has equal probability (one half) to be brown and to be blue. As a result the child's first allele has probability $\frac{5}{12}$ to be blue and his second allele has the probability $\frac{1}{2}$. He has blue eye if both of these are blue, that is in $\frac{5}{24}$ cases. In the remaining $\frac{19}{24}$ he has brown eyes.

Given a comprimed family tree, determine the probability the unborn child has brown eyes. There are at most 10^6 nodes in the tree and the weights are at most 10^{12}.