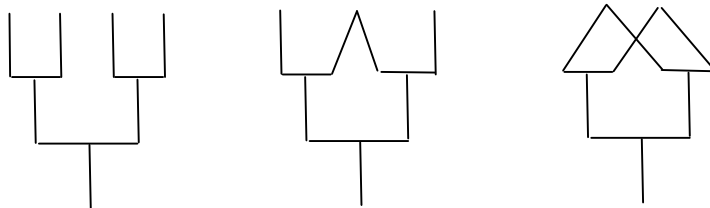


Supervisor: Jotun Hein

Project Title: Counting Pedigrees up to Isomorphism

Evolutionary Biology is dominated by relationship of species, individuals, sequences and other “entities” that can be traced back to common ancestor(s) by some copying and modification mechanism. For species this will typically lead to the phylogeny, that in many cases is a tree with a root, where the leaves are labelled but the internal nodes are not. If the length of the branches are ignored, biologist talk of tree topologies, which are discrete object and can be counted as a function of the number of leaves. A famous result going back to Ernst Schröder 1870 is that if internal nodes correspond to duplications (not triplications and higher) then the number of tree topologies with k leaves is $1 \cdot 3 \cdot 5 \dots (2k-5)$. The concept of phylogenies stretches back only a few centuries, while pedigrees (the concepts of our individual ancestors) goes back millennia. Counting problems relating to number of pedigrees haven’t been studied as much and contains a series of open problems. One reason is that it is probably much harder and will also have a much higher asymptotic growth. Knowing this number is of potential practical interest as a measure of how hard it is to determine which pedigree describe the relationship of a certain number of generations back in time for a certain number of extant individuals. To define and limit the project, it will be assumed that we operate with discrete time and generations and since a pedigree in principle stretches arbitrarily far back in time, we will assume we only register T generations back starting with the present generation as 0. Let i be the number of extant individuals we have sampled. The population could be finite – N – and the number of ancestors could then not be more than this number at any given time. The pedigree could be sex labelled or not. If the pedigree isn’t sex labelled, it must for obvious reasons to be possible to sex label it. (We ignore the complication of identical twins and cloning). The simplest possible problem would be to sample one individual, count sex labelled pedigrees, let the population be infinite and for instance observe the ancestry 2 generations back in time: In this case there would be 3 possible pedigrees: the single individual must have 2 distinct parents, the mother and father could have 0,1 or 2 parents in common



If they were sex labelled, the first (“normal”) pedigree could either have a male or a female in the present (i.e. 2 sex labellings), the second would additionally imply a choice for the grand parent with 2 offspring (i.e. 4 sex labellings) and the last would again have 2 sex labellings.

Project: The project will start by reviewing a few key articles/chapters to lay out the problem description. The project can start out in a very sure way, as it isn’t difficult to write a recursion that counts the number of pedigrees T generations back in time, given the pedigrees $T-1$ generations back in time lumped in a series of groups. This recursion is similar to recursions developed by Song, Lyngsø and Hein (2005). This would allow for counting 10-15 generations back in time. Going beyond this, would require combinatorial insights. The project can be extended by considering sex labelled/unlabelled, constant population size and possibly relaxing the assumptions of constant generation time.

Characterize or find inequalities relating to the asymptotic growth of the number of pedigrees as a function of extant individuals and number of generations back in time could be a possibility in case of solid progress.

References

- Chang, J. (1999) "Recent Common Ancestor of All Present Human Individuals" Adv. Appl. Prob. vol. 31.4
- Derrida et al. (2000) "On the genealogy of a population of biparental individuals" J. Theor. Biol. 203.303-315
- Hein, J. J. (2004) "Pedigrees for all Humanity" Nature 431.512-13.

Hein, Schierup and Wiuf (2005) "Gene Genealogies, Variation and Evolution" Oxford University Press
Rhode, DLT, S. Olson & J. Chang (2004) "Modelling the recent common ancestry of all living humans" Nature 431.562-
Semple and Steel (2003) "Phylogenetics" Oxford University Press
Song, Y., Lyngsø, R. and J Hein (2005 submitted) "Counting ancestral states in population genetics" Journal of Bioinformatics and
Computational Biology.
Steel and Hein (2005) "Reconstructing Pedigrees and a Tree of Life: A Combinatorial Perspective" manuscript
Thomas, A. and Cannings (2004) "Simulating realistic zero loop pedigrees using bipartite Prüfer code and graphical modelling"
Mathematical Medicine and Biology 21.335-345