

# Non-Identifiable Pedigrees and a Bayesian Solution

B. Kirkpatrick

University of British Columbia

**Abstract.** Some methods aim to correct or test for relationships or to reconstruct the pedigree, or family tree. We show that these methods cannot resolve ties for correct relationships due to identifiability of the pedigree likelihood which is the probability of inheriting the data under the pedigree model. This means that no likelihood-based method can produce a correct pedigree inference with high probability. This lack of reliability is critical both for health and forensics applications.

Pedigree inference methods use a structured machine learning approach where the objective is to find the pedigree graph that maximizes the likelihood. Known pedigrees are useful for both association and linkage analysis which aim to find the regions of the genome that are associated with the presence and absence of a particular disease. This means that errors in pedigree prediction have dramatic effects on downstream analysis.

In this paper we present the first discussion of multiple typed individuals in non-isomorphic pedigrees,  $\mathcal{P}$  and  $\mathcal{Q}$ , where the likelihoods are non-identifiable,  $Pr[G | \mathcal{P}, \theta] = Pr[G | \mathcal{Q}, \theta]$ , for all input data  $G$  and all recombination rate parameters  $\theta$ . While there were previously known non-identifiable pairs, we give an example having data for multiple individuals.

Additionally, deeper understanding of the general discrete structures driving these non-identifiability examples has been provided, as well as results to guide algorithms that wish to examine only identifiable pedigrees. This paper introduces a general criteria for establishing whether a pair of pedigrees is non-identifiable and two easy-to-compute criteria guaranteeing identifiability. Finally, we suggest a method for dealing with non-identifiable likelihoods: use Bayes rule to obtain the posterior from the likelihood and prior. We propose a prior guaranteeing that the posterior distinguishes all pairs of pedigrees.

---

**Keywords:** pedigree genetics, discrete probability, identifiability.