Forensic Science International: Genetics 29 (2017) 218–224

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Exact likelihood ratio calculations for pairwise cases

Thore Egeland [a,*], Nádia Pinto [b,c,d], António Amorim [b,c,e]

[a] Norwegian University of Life Sciences, Norway
[b] Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal
[c] Instituto de Investigação e Inovação em Saúde (i3s), Universidade do Porto, Porto, Portugal
[d] Centro de Matemática da Universidade do Porto, Porto, Portugal
[e] Faculdade de Ciências, Universidade do Porto, Porto, Portugal

ARTICLE INFO

Article history:
Received 18 October 2016
Received in revised form 22 March 2017
Accepted 24 April 2017
Available online 27 April 2017

ABSTRACT

Some practical and theoretical aspects of evaluation of evidence based on the likelihood ratio (LR) in kinship cases are discussed. If relationships are complex or if complicating factors like mutation, correction for population structure or silent alleles need to be accounted for, available software may fail. We present an explicit general formula for non-inbred pairwise cases. Equipped with this formula it is possible to evaluate, say, how strongly a shared rare allele, points towards a specific relationship. Moreover, a general expression as the one presented, adds to the understanding of models and the underlying biological mechanisms. It is also useful for checking software and defining the limitations of programs. Some ideas for improving software may also be generated by the derivation of exact expressions.

We argue that a *proportional* mutation model is well suited from a pragmatic point of view and derive some theoretical properties of this model. Several examples based on the general pairwise formula and its implementation in the freely available `R` package `mut` are presented.

1. Introduction

The broad topic for this paper is likelihood ratio (LR) calculations in kinship cases. For most practical problems faced by caseworkers, there is software available that can calculate the required LRs. There are, however, some exceptions motivating this paper. We restrict attention to problems involving two individuals, so–called pairwise cases. This is a relevant class of problems to study for several reasons and lends itself to exact expressions. Our focus is beyond the standard applications and typically involves one or more of the following complicating factors: (i) large pedigrees, (ii) mutations, recently discussed in [10], (iii) population stratification [1], and (iv) silent alleles. Available software including `Familias` [7,3] fail to calculate LR for large pedigrees with *marriage loops*. One example is shown in Fig. 1. Two remotely related individuals share a rare allele denoted 1. This is a case of some independent interest as it addresses the question: What is the statistical evidence from a rare variant? As there is a large number of meioses, it is hard to justify not modelling mutations. In addition, the frequency of the allele or haplotype may be in the order of magnitude of the mutation rate.

We present a formula for general pairwise cases accomodating all mentioned complicating factors as well as freely available software. This formula is also relevant as it may be used to check software and precisely define the limitations of currently available implementations. In addition, with a general formula, it is possible to estimate parameters optimally using the maximum likelihood method and also better understand how parameters describing, say, mutations and population stratification influence results.

2. Methods

We consider general non-inbred pairwise cases as described by the IBD (Identical By Descent) parameter $\kappa = (\kappa_0, \kappa_1, \kappa_2)$. An allele in one individual is IBD to an allele in another individual if it comes from the same ancestral allele within the pedigree. The probability that two individuals share i alleles IBD is κ_i. Obviously $\kappa_0 + \kappa_1 + \kappa_2 = 1$ in addition to $\kappa_1^2 \geq 4\kappa_0\kappa_2$ proved in [13]. We first review some useful formulae for the LRs for pairwise relationships. Thereafter these results are generalized.

* Corresponding author at: Norwegian University of Life Sciences, 1432 Aas, Norway.
E-mail address: Thore.Egeland@nmbu.no (T. Egeland).

http://dx.doi.org/10.1016/j.fsigen.2017.04.018
1872-4973/

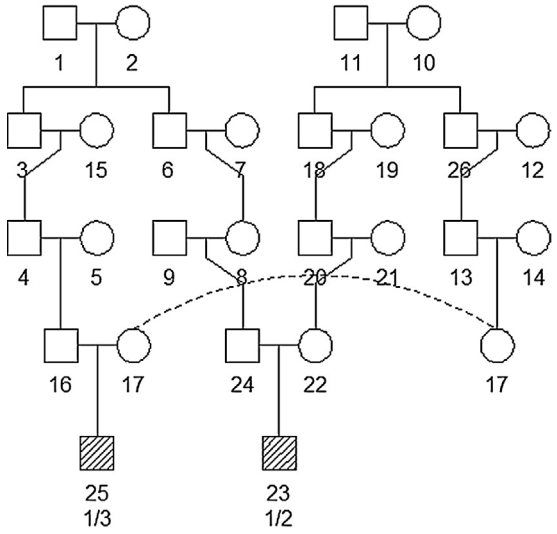

Fig. 1. The boys 23 and 25 are double third cousins. There is a marriage loop: the second cousins 16 and 24 have children with respectively the second cousins 17 and 22. The pedigree is included as an example which is beyond the reach of forensic software we know, but which computes easily with the approach of the present paper also when mutation and theta correction are accounted for.

2.1. Mutations not possible

As shown in [14, p. 42], the likelihood function for one marker can be written

$$L(\kappa) = \kappa_0 P(G|I=0) + \kappa_1 P(G|I=1) + \kappa_2 P(G|I=2) \tag{2.1}$$

where $G := (g_1, g_2)$ are the genotypes. I denotes the number of IBD alleles and 0, 1 and 2 correspond to "unrelated","parent offspring", and "monozygotic twins" respectively. The dependence on g_1 and g_2 and the corresponding frequencies p_{g_1} and p_{g_2} is omitted in the notation $L(\kappa)$.

The LR for relatedness with IBD coefficients (κ_0, κ_1, κ_2) compared to unrelated is presented in [11] and can be written

$$\begin{aligned} LR(G) &= \kappa_0 + \kappa_1 \frac{P(G|I=1)}{P(G|I=0)} + \kappa_2 \frac{P(G|I=2)}{P(G|I=0)} \\ &= \kappa_0 + \kappa_1 PI(G) + \kappa_2 \mathbb{1}(g_1 = g_2) \frac{1}{p_{g_1}} \end{aligned} \tag{2.2}$$

using (2.1) where PI abbreviates paternity index, and $\mathbb{1}(g_1 = g_2) = 1$ if $g_1 = g_2$ and 0 otherwise.

2.2. General paternity case

We next discuss how mutation, theta correction and silent alleles are all taken care of in a general formula for the PI.

2.2.1. Mutation

The paternity index for a parent with genotype $g_1 = a/b$ and a child with genotype $g_2 = c/d$, where the alleles may or may not differ, equals [9]

$$PI(g_1, g_2, M) = \frac{2^{-\mathbb{1}(a=b)-\mathbb{1}(c=d)} p_a p_b ((m_{ac} + m_{bc}) p_d + (m_{ad} + m_{bd}) p_c)}{2^{-\mathbb{1}(a=b)-\mathbb{1}(c=d)} 4 p_a p_b p_c p_d} \tag{2.3}$$

$$= \frac{1}{4} \frac{(m_{ac} + m_{bc}) p_d + (m_{ad} + m_{bd}) p_c}{p_c p_d}. \tag{2.4}$$

Here m_{ij} denotes the probability that allele i ends up as j. With 0 mutation rate, $m_{ij} = \mathbb{1}(i = j)$. The first expression is from the derivation of (2.4) in [3, pp. 173–174]. The numerator and denominator of (2.3) are the likelihoods for the two hypotheses and they will be needed below. Parent-child relationships are examples of pairwise cases and are referred to as "duo cases" below.

Example 2.1. Consider a duo case and assume mutations are not possible. Without loss of generality we can consider a marker with four alleles denoted 1, 2, 3, 4 and so for instance $m_{11} = 1$ while $m_{12} = 0$. Consider first three paternity examples. If the individuals are heterozygous sharing no, one or both alleles Eq. (2.4) gives

$$\begin{aligned} &PI(1/2, 3/4) = 0, \quad PI(1/2, 1/3) = \frac{1}{4p_1}, \\ &PI(1/2, 1/2) = \frac{p_1 + p_2}{4 p_1 p_2}. \end{aligned} \tag{2.5}$$

Likelihood ratios comparing any non-inbred pairwise relationship to unrelated for the same genotype data are found by inserting the paternity indices into (2.2) as exemplified below for full–sibs ($\kappa_0 = 0.25, \kappa_1 = 0.5, \kappa_2 = 0.25$). The numerical values are for $p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.3$ and $p_4 = 0.4$.

$$\begin{aligned} LR(1/2, 3/4) &= 0.25 + 0.5 \times 0 + 0.25 \times 0 \times \frac{1}{2 p_1 p_2} = 0.25, \\ LR(1/2, 1/3) &= 0.25 + 0.5 \times \frac{1}{4 p_1} + 0.25 \times 0 \times \frac{1}{2 p_1 p_2} = 1.5, \\ LR(1/2, 1/2) &= 0.25 + 0.5 \times \frac{p_1 + p_2}{4 p_1 p_2} + 0.25 \times 1 \times \frac{1}{2 p_1 p_2} = 8.375. \end{aligned}$$

Example 2.2. A father and a son are missing and biological samples A and B have been genotyped from two bodies. If this is a parent child relationship, we do not know if A is from the father and B from the son or the other way around. We consider the hypotheses

H_1: The bodies correspond to individuals related as father and son.
H_2: The bodies correspond to unrelated individuals.
Consider, for simplicity, a SNP marker with alleles 1 and 2, and frequencies p_1 and $p_2 = 1 - p_1$, for which we observe 1/1 in sample A and 2/2 in B. With the simplifying assumptions (no genotyping errors, no drop-out, no silent alleles, no mutations) the likelihood ratio for this marker would be 0. If, on the other hand we allow for mutations, the LR will be greater than 0 and the overall LR based on all markers may exceed stipulated thresholds for declaring paternity. Assume there is a probability m_{12} of a mutation from 1 to 2 while m_{21} is the probability from 2 to 1. From (2.4) follows $LR_{AB} = m_{12}/p_2$ if A is the father of B, while if B is the father of A, $LR_{BA} = m_{21}/p_1$ and thus the likelihoods ratios may differ substantially. For instance if $m_{12} = m_{21} = R$, $p_1 = p$, $p_2 = 1 - p$, we find that the ratio of the LRs is

$$\frac{LR_{AB}}{LR_{BA}} = \frac{p}{1-p}$$

which equals 1 only when $p = 0.5$. If for instance, $p = 0.8$, the ratio is 4 and in other words the likelihood ratio is 4 times larger if we assume A to be the father of B compared to the alternative. In cases where there is no information on the order of A and B, we have encountered a practical problem. In this paper we describe a practical *mathematical* solution to the problem: We use a mutation model that guarantees

$$LR_{AB} = LR_{BA}, \text{ i.e., the } LR \text{ is unchanged if genotypes are swapped.} \tag{2.6}$$

Throughout we refer to the latter equality as the *swap property*. Observe that (2.6) above is equivalent to

$$p_1 m_{12} = p_2 m_{21}$$