

Testing for kinship

In Chapter 4, we discussed a number of paternity problems such as paternity testing for the with-mother and without-mother cases, incest cases and determinations of both parents, etc. We have also considered situations in which relatives of the alleged father are involved, for example the defense proposition that a relative of the alleged father is the true father of the child, and when the alleged father is not available but his relative is. All of the above analyses are concerned about the determination of the ‘father and child’ relationship. In this particular chapter, we are going to investigate more general relationships between two persons (Fung *et al.* 2003a; Li and Sacks 1954) and those among three persons (Fung *et al.* 2005, 2006). Besides, unlike in Chapter 4, in which Hardy–Weinberg equilibrium (HWE) is assumed, the relationships of the persons involved here are determined under the situation that they belong to a subdivided or structured population. More complex paternity and kinship problems (Fung 2003b) with Hardy–Weinberg equilibrium are also investigated.

5.1 Kinship testing of any two persons: HWE

In addition to the parent–child determination in traditional parentage testing (Fung *et al.* 2002; Lee *et al.* 2000, 1999; Thomson *et al.* 1999), other kinds of relationships between individuals also need to be tested in practice. For example, Thomson *et al.* (2001) analyzed sibling relationships using STR loci; Gaytmenn *et al.* (2002) studied the sensitivity and specificity of sibship calculations. The use and abuse of the full sibling and half sibling indices in immigration cases were discussed by Gorlin and Polesky (2000).

We would like to determine a specific relationship between two typed persons. The following propositions are of interest:

$$\begin{aligned} H_p: & \text{the two persons are biologically related;} \\ H_d: & \text{the two persons are biologically unrelated.} \end{aligned} \tag{5.1}$$

The relationship between the two persons can be of various sorts, for example parent–offspring, uncle–nephew, half siblings, etc. In this particular section, Hardy–Weinberg equilibrium is assumed. Let the genotypes of the two persons be denoted by Y and Z , respectively. For

Table 5.1 The joint genotype probabilities $P(Y, Z)$ for all Y and Z combinations when the population is in Hardy–Weinberg equilibrium.

Y	Z	$P(Y, Z)$
$A_i A_i$	$A_i A_i$	$k_0 p_i^4 + 2k_1 p_i^3 + k_2 p_i^2$
$A_i A_i$	$A_i A_j$	$2k_0 p_i^3 p_j + 2k_1 p_i^2 p_j$
$A_i A_i$	$A_j A_j$	$k_0 p_i^2 p_j^2$
$A_i A_i$	$A_j A_k$	$2k_0 p_i^2 p_j p_k$
$A_i A_j$	$A_i A_j$	$4k_0 p_i^2 p_j^2 + 2k_1 p_i^2 p_j + 2k_1 p_i p_j^2 + 2k_2 p_i p_j$
$A_i A_j$	$A_i A_k$	$4k_0 p_i^2 p_j p_k + 2k_1 p_i p_j p_k$
$A_i A_j$	$A_k A_l$	$4k_0 p_i p_j p_k p_l$

simplicity, we use the following notation to describe the relationship between the two persons in the hypotheses. The pair of hypotheses in (5.1) is then formulated as

$$\begin{aligned} H_p &: (Y, Z) \sim (k_0, 2k_1, k_2); \\ H_d &: (Y, Z) \sim (1, 0, 0), \end{aligned} \quad (5.2)$$

where $(k_0, 2k_1, k_2)$ are the relatedness coefficients [see Equation (3.18)] of the two persons with common values such as $(0, 1, 0)$ for parent–offspring, $(0.5, 0.5, 0)$ for half siblings and $(1, 0, 0)$ for unrelated persons, etc. The joint genotype probabilities $P(Y, Z)$ for all possible genotype combinations of Y and Z are listed in Table 5.1 when the population is in Hardy–Weinberg equilibrium. The results in Table 5.1 can be derived using a general formula given in Equation (7.21).

The likelihood ratio about the hypothesis pair in (5.1) is

$$\begin{aligned} LR &= \frac{P(\text{evidence}|H_p)}{P(\text{evidence}|H_d)} \\ &= \frac{P(Y, Z|H_p)}{P(Y, Z|H_d)} \\ &= \frac{P(Z|Y, H_p)}{P(Z|Y, H_d)} \frac{P(Y|H_p)}{P(Y|H_d)} \\ &= \frac{P(Z|Y, H_p)}{P(Z|Y, H_d)}. \end{aligned} \quad (5.3)$$

Suppose that the two persons have genotypes $Y = A_i A_j$ and $Z = A_i A_k$ at a particular locus. The numerator of the likelihood ratio in Equation (5.3), based on the law of total probability in Equation (2.19), can be evaluated as

$$\begin{aligned} P(Z = A_i A_k | Y = A_i A_j, H_p) &= \sum_{t=0}^2 P(Z = A_i A_k | Y = A_i A_j, H_p, \text{ the two persons have } t \text{ ibd alleles}) \\ &\quad \times P(\text{the two persons have } t \text{ ibd alleles} | Y = A_i A_j, H_p) \\ &= (2p_i p_k)k_0 + (p_k \times 1/2)2k_1 + 0 \times k_2. \end{aligned}$$

The second term ($p_k \times 1/2$) is explained as follows: given $Y = A_i A_j$ and the two persons have one ibd allele, the probability of the ibd allele being A_i is $1/2$ and the probability of the remaining allele being A_k is p_k . Hence,

$$P(Z = A_i A_k | Y = A_i A_j, H_p) = 2p_i p_k k_0 + p_k k_1.$$

The denominator of the likelihood ratio is simply

$$P(Z = A_i A_k | Y = A_i A_j, H_d) = 2p_i p_k,$$

since, under H_d that the two persons are unrelated, the probability that the genotype $Z = A_i A_k$ is independent of the genotype $Y = A_i A_j$. Thus, the likelihood ratio corresponding to the set of hypotheses (5.1) or (5.2) on kinship determination is

$$\begin{aligned} LR &= (2p_i p_k k_0 + p_k k_1) / (2p_i p_k) \\ &= k_0 + k_1 / (2p_i). \end{aligned}$$

The likelihood ratios for other combinations of genotypes Y and Z can be derived in similar ways. These ratios are all listed in Table 5.2, which can also be derived based on Table 5.1 [see also Li and Sacks (1954)].

The likelihood ratios in Table 5.2 correspond to the hypothesis pair (5.1) in which the relatedness coefficients ($k_0, 2k_1, k_2$) are used to describe the relationship between the two persons under H_p . When the coefficients take values (0, 1, 0) for father–child, the likelihood ratios given in Table 5.2 reduce to the paternity indices (PI 's), as reported in Table 4.5 for paternity testing in a motherless case.

Sometimes, the defense proposition H_d is that the two persons are not unrelated: consider the pedigree diagram in Figure 5.1 in which persons 5 and 6 are couple, and it is also known that person 5 is the biological mother of person 8. Suppose that there is a query if in fact person 4 instead of person 6 is the biological father of person 8. Suppose persons 4, 5 and 6 are unavailable, and only persons 7 and 8 (see diagram) are available for typing; in this case, we have the following propositions which describe whether persons 7 and 8 are related as half siblings or first cousins, i.e.

$$\begin{aligned} H_p: (Y, Z) &\sim (0.5, 0.5, 0); \\ H_d^*: (Y, Z) &\sim (0.75, 0.25, 0). \end{aligned} \quad (5.4)$$

Table 5.2 The likelihood ratios about two competing hypotheses $H_p : (Y, Z) \sim (k_0, 2k_1, k_2)$ versus $H_d : (Y, Z) \sim (1, 0, 0)$.

Y	Z	Likelihood ratio
$A_i A_i$	$A_i A_i$	$k_0 + 2k_1/p_i + k_2/p_i^2$
$A_i A_i$	$A_i A_j$	$k_0 + k_1/p_i$
$A_i A_i$	$A_j A_j$	k_0
$A_i A_i$	$A_j A_k$	k_0
$A_i A_j$	$A_i A_j$	$k_0 + k_1(p_i + p_j)/(2p_i p_j) + k_2/(2p_i p_j)$
$A_i A_j$	$A_i A_k$	$k_0 + k_1/(2p_i)$
$A_i A_j$	$A_k A_l$	k_0

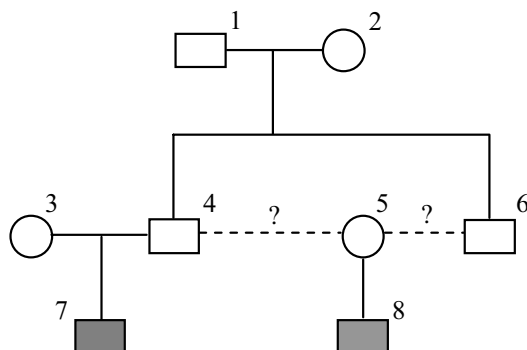


Figure 5.1 A pedigree diagram with only persons 7 and 8 available for typing. Interested in testing whether persons 7 and 8 are related as half siblings or first cousins.

The associated likelihood ratios can be obtained in the following way.

We consider the hypotheses with relationship half siblings versus unrelated:

$$\begin{aligned} H_p: (Y, Z) &\sim (0.5, 0.5, 0); \\ H_d: (Y, Z) &\sim (0, 0, 1), \end{aligned} \quad (5.5)$$

and relationship first cousins versus unrelated:

$$\begin{aligned} H_d^*: (Y, Z) &\sim (0.75, 0.25, 0); \\ H_d: (Y, Z) &\sim (0, 0, 1). \end{aligned} \quad (5.6)$$

Suppose that the associated likelihood ratios are, respectively, LR_1 and LR_2 , which can be evaluated from Table 5.2. It is obvious that the likelihood ratio for hypotheses set (5.4) can be obtained as

$$LR = LR_1/LR_2. \quad (5.7)$$

We consider the genotypes of the two persons (Table 5.3) and are interested in testing whether they are half siblings or first cousins; the pair of hypotheses is as given in (5.4). We first evaluate the likelihood ratios for the hypothesis pair (5.5):

$$\begin{aligned} \text{D3S1358} : LR_1 &= k_0 + k_1(p_{15} + p_{17})/(2p_{15}p_{17}) + 0 \\ &= 0.5 + 0.25 \times (0.331 + 0.239)/(2 \times 0.331 \times 0.239) = 1.401, \end{aligned}$$

$$\text{vWA} : LR_1 = k_0 + k_1/(2p_{15}) = 0.5 + 0.25/(2 \times 0.035) = 4.071,$$

$$\text{FGA} : LR_1 = k_0 + k_1/p_{22} = 0.5 + 0.25/0.178 = 1.904.$$

Table 5.3 Genotype data of two persons Y and Z at loci D3S1358, vWA and FGA.

Locus	Y	Z
D3S1358	15/17	15/17
vWA	14/15	15/19
FGA	22/22	22/23

Then, we assess the likelihood ratios for the hypothesis pair (5.6) in a similar way. They are

$$\text{D3S1358} : LR_2 = 0.75 + 0.125 \times (0.331 + 0.239)/(2 \times 0.331 \times 0.239) = 1.2,$$

$$\text{vWA} : LR_2 = 0.75 + 0.125/(2 \times 0.035) = 2.536,$$

$$\text{FGA} : LR_2 = 0.75 + 0.125/0.178 = 1.452.$$

Thus, the likelihood ratios for the pair of hypotheses (5.4) at the three loci, based on Equation (5.7), are, respectively, $1.401/1.2 = 1.17$, $4.071/2.536 = 1.61$, and $1.904/1.452 = 1.31$. The overall likelihood ratio is then $1.17 \times 1.61 \times 1.31 = 2.47$, which seems to provide a larger support to the hypothesis that the two persons are half siblings related.

5.2 Computer software

A computer software has been developed to deal with various two-person kinship problems. The software is named EasyDNA_2Persons which consists of the following steps:

Steps in running the EasyDNA 2Persons software

- 1 Click the *Load frequency file* button after loading the EasyDNA program, then select the appropriate file
- 2 Choose the allele pairs at the locus for Y and Z
- 3a Choose the appropriate relation between Y and Z under H_p (which is, for the above example, *Half siblings*)
- 3b Choose the appropriate relation between Y and Z under H_d (which is, for the above example, *First cousins*)
- 4 Click the *Calculate* button
- 5 Repeat steps 2 and 4 for each of the remaining loci; step 3 (3a–3b) is blocked, since it is no longer needed for the remaining loci.

The procedure steps are straightforward and readers can easily get familiar with the running of the software. It is noted that the theory provided in Section 5.1 and the associated computer software can be used to determine the relationship between any two persons. For illustration, we analyze the example given in Table 5.3 using the EasyDNA_2Persons software. Figure 5.2 presents the captured screen in the running of the software. The likelihood ratios for individual loci are obtained as 1.17, 1.61 and 1.31, with an overall likelihood ratio of 2.47. These values are the same as those obtained above by formulas.

5.3 Kinship testing of two persons: subdivided populations

5.3.1 Joint genotype probability

The issue of population subdivision for paternity and kinship determination has been addressed. Balding and Nichols (1995) considered paternity testing for the case in which the mother, alleged father and ‘alternative father’ all belong to the same subpopulation. Ayres (2000) proposed tests for kinships in subdivided/structured populations. Clayton *et al.* (2002)

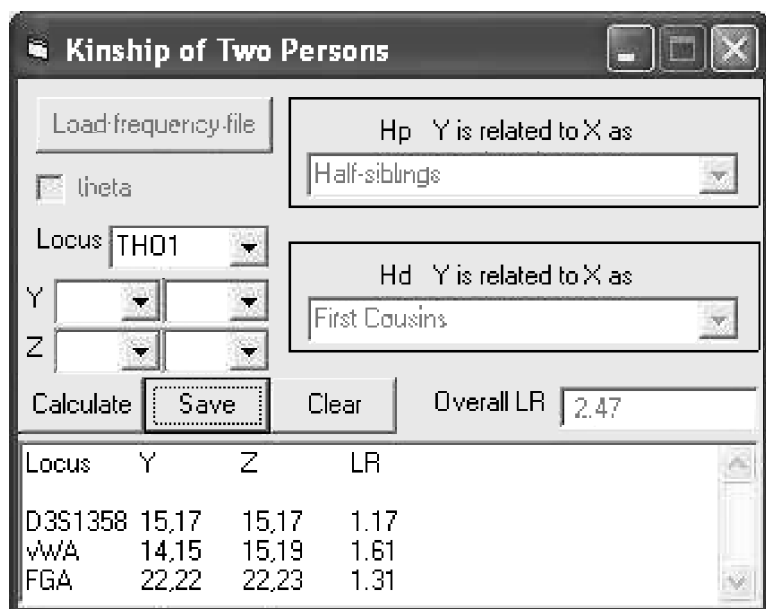


Figure 5.2 Captured screen for running the EasyDNA_2Persons software for determination of kinship for two persons Y and Z whose genotypes are provided in Table 5.3.

discussed that it might make more sense to take account of population subdivision in paternity testing and they referred to the results of Ayres (2000).

In this section, we are using the conditional probability formula for a subdivided population as given in Equation (3.17), where θ there measures the degree of subdivision. We shall derive expressions of the joint genotype probabilities for kinship of any two persons in a subdivided population and likelihood ratios for testing kinship of the two individuals (Fung *et al.* 2003a).

In order to test whether two given persons have the specific relationship claimed, we first need to find the joint genotype probability $P(Y, Z)$, where Y and Z are respectively genotypes of the two persons who belong to the same subdivided population. Denote respectively the paternal and maternal alleles of Y and Z as Y_P, Z_P, Y_M and Z_M , then the relatedness coefficients $(k_0, 2k_1, k_2)$ can be expressed as

$$k_2 = P(Y_P \equiv Z_P, Y_M \equiv Z_M) + P(Y_P \equiv Z_M, Y_M \equiv Z_P),$$

$$2k_1 = P(Y_P \equiv Z_P) + P(Y_P \equiv Z_M) + P(Y_M \equiv Z_P) + P(Y_M \equiv Z_M),$$

and

$$k_0 = P(\text{no ibd alleles}) = 1 - 2k_1 - k_2,$$

where the symbol ' $\equiv$ ' denotes the ibd relationship of alleles (Evetts and Weir 1998).

It is obvious that the relatedness coefficients will play an important role in the evaluation of joint genotype probability $P(Y, Z)$. For two related individuals described with relatedness coefficients $(k_0, 2k_1, k_2)$, there are seven possible combinations of the alleles of Y and Z (irrespective of order). In the following, we consider the simple case in which $Y = Z = A_i A_i$

and demonstrate the principle for the evaluation of their joint genotype probability $P(Y, Z)$. In this case, we have $Y_P = Y_M = Z_P = Z_M = A_i$, so

$$\begin{aligned}
 & P(Y = A_i A_i, Z = A_i A_i) \\
 &= P(Y_P = Y_M = Z_P = Z_M = A_i) \\
 &= \sum_{t=0}^2 P(Y_P = Y_M = Z_P = Z_M = A_i, t \text{ ibd alleles}) \\
 &= P(Y_P = Y_M = Z_P = Z_M = A_i, \text{no ibd alleles}) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_P \equiv Z_P) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_P \equiv Z_M) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_M \equiv Z_P) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_M \equiv Z_M) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_P \equiv Z_P, Y_M \equiv Z_M) \\
 &\quad + P(Y_P = Y_M = Z_P = Z_M = A_i, Y_P \equiv Z_M, Y_M \equiv Z_P) \\
 &= P(\text{no ibd alleles})P(A_i, A_i, A_i, A_i) \\
 &\quad + [P(Y_P \equiv Z_P) + P(Y_P \equiv Z_M) + P(Y_M \equiv Z_P) + P(Y_M \equiv Z_M)]P(A_i, A_i, A_i) \\
 &\quad + [P(Y_P \equiv Z_P, Y_M \equiv Z_M) + P(Y_P \equiv Z_M, Y_M \equiv Z_P)]P(A_i, A_i).
 \end{aligned}$$

This gives immediately the corresponding result reported in the first genotype combination in Table 5.4. The joint genotype probabilities for the other six possible genotype combinations can be derived similarly and we omit the details.

Note that the results in Table 5.4 are general results that can be applied in various ways for the evaluation of joint genotype probabilities. For example, in the penultimate row, under Hardy–Weinberg equilibrium, the probabilities $P(A_i, A_i, A_j, A_k)$ and $P(A_i, A_j, A_k)$ are evaluated as $p_i p_i p_j p_k$ and $p_i p_j p_k$, respectively. Thus, the joint genotype probability $P(Y = A_i A_j, Z = A_i A_k)$ is obtained as $4k_0 p_i^2 p_j p_k + 2k_1 p_i p_j p_k$ under Hardy–Weinberg equilibrium.

In a subdivided population in which Hardy–Weinberg equilibrium does not hold, the evaluation of probabilities is implemented by employing the recursive formula in Equation (3.17). For example, in row 4 of Table 5.4,

$$\begin{aligned}
 P(A_i, A_i, A_j, A_k) &= P(A_i)P(A_i|A_i)P(A_j|A_i, A_i)P(A_k|A_i, A_i, A_j) \\
 &= \left[\frac{(1-\theta)p_i}{1+(0-1)\theta} \right] \left[\frac{\theta+(1-\theta)p_i}{1+(1-1)\theta} \right] \left[\frac{(1-\theta)p_j}{1+(2-1)\theta} \right] \left[\frac{(1-\theta)p_k}{1+(3-1)\theta} \right] \\
 &= \frac{(1-\theta)^2 p_i p_j p_k [\theta+(1-\theta)p_i]}{(1+\theta)(1+2\theta)}
 \end{aligned}$$

and the genotype probability $P(Y, Z)$ can be obtained accordingly.

Table 5.4 The joint genotype probabilities $P(Y, Z)$ for all Y and Z combinations where Y and Z come from the same subdivided population, from Fung *et al.* (2003a). (Reproduced by permission of Elsevier.)

Y	Z	$P(Y, Z)$
$A_i A_i$	$A_i A_i$	$k_0 P(A_i, A_i, A_i, A_i) + 2k_1 P(A_i, A_i, A_i) + k_2 P(A_i, A_i)$
$A_i A_i$	$A_i A_j$	$2k_0 P(A_i, A_i, A_i, A_j) + 2k_1 P(A_i, A_i, A_j)$
$A_i A_i$	$A_j A_j$	$k_0 P(A_i, A_i, A_j, A_j)$
$A_i A_i$	$A_j A_k$	$2k_0 P(A_i, A_i, A_j, A_k)$
$A_i A_j$	$A_i A_j$	$4k_0 P(A_i, A_i, A_j, A_j) + 2k_1 P(A_i, A_i, A_j) + 2k_1 P(A_i, A_j, A_j) + 2k_2 P(A_i, A_j)$
$A_i A_j$	$A_i A_k$	$4k_0 P(A_i, A_i, A_j, A_k) + 2k_1 P(A_i, A_j, A_k)$
$A_i A_j$	$A_k A_l$	$4k_0 P(A_i, A_j, A_k, A_l)$

One nice feature of the probabilities presented in Table 5.4 is that they can be applied for testing for kinship of any two persons in a subdivided population. Consider the same set of hypotheses as given in (5.2):

$$H_p : (Y, Z) \sim (k_0, 2k_1, k_2);$$

$$H_d : (Y, Z) \sim (1, 0, 0).$$

The genotype probabilities $P(Y, Z|H_p)$ and $P(Y, Z|H_d)$ are straightforward from Table 5.4 and their ratios, i.e. the likelihood ratios for all seven genotype combinations of Y and Z , are listed in Table 5.5. For example in the second row, where $Y = A_i A_i$ and $Z = A_i A_j$, we have from Table 5.4 that

$$\begin{aligned}
 LR &= \frac{2k_0 P(A_i, A_i, A_i, A_j) + 2k_1 P(A_i, A_i, A_j)}{2P(A_i, A_i, A_i, A_j)} \\
 &= k_0 + \frac{k_1}{P(A_i|A_i, A_i, A_j)} \\
 &= k_0 + \frac{k_1(1 + 2\theta)}{2\theta + (1 - \theta)p_i}.
 \end{aligned}$$

The likelihood ratios for the other six cases can be obtained similarly.

For the usual paternity testing in the no-mother case, the two competing hypotheses are

$$H_p : (C, AF) \sim (0, 1, 0);$$

$$H_d : (C, AF) \sim (1, 0, 0).$$

The paternity indices (PI 's) can be obtained from Table 5.5, with specific values of $k_0 = 0$ and $2k_1 = 1$. These indices are reported in Table 5.6. They can be used when the alleged father and the child belong to a subdivided population with the degree of subdivision θ . When the population is in Hardy–Weinberg equilibrium, i.e. $\theta = 0$, the PI 's in Table 5.6 reduce to those reported in Table 4.5.

Table 5.5 The likelihood ratios about two competing hypotheses $H_p : (Y, Z) \sim (k_0, 2k_1, k_2)$ versus $H_d : (Y, Z) \sim (1, 0, 0)$ in a subdivided population, from Fung *et al.* (2003a). (Reproduced by permission of Elsevier.)

Y	Z	Likelihood ratio
$A_i A_i$	$A_i A_i$	$k_0 + \frac{2k_1(1+2\theta)}{3\theta + (1-\theta)p_i} + \frac{k_2(1+\theta)(1+2\theta)}{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}$
$A_i A_i$	$A_i A_j$	$k_0 + \frac{k_1(1+2\theta)}{2\theta + (1-\theta)p_i}$
$A_i A_i$	$A_j A_j$	k_0
$A_i A_i$	$A_j A_k$	k_0
$A_i A_j$	$A_i A_j$	$k_0 + \frac{k_1(1+2\theta)[2\theta + (1-\theta)(p_i + p_j)] + k_2(1+\theta)(1+2\theta)}{2[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}$
$A_i A_j$	$A_i A_k$	$k_0 + \frac{k_1(1+2\theta)}{2[\theta + (1-\theta)p_i]}$
$A_i A_j$	$A_k A_l$	k_0

5.3.2 Relatives involved

In fact, Table 5.5 also provides the likelihood ratio for any two propositions H_p and H_d , which is simply the ratio of the likelihood ratio about H_p versus unrelated and the likelihood ratio about H_d versus unrelated. Particularly, we consider the paternity testing in the no-mother

Table 5.6 Paternity index (PI) for the competing hypotheses H_p : AF is true father of the child versus H_d : the father is a random unrelated man, i.e. $H_p : (C, AF) \sim (0, 1, 0)$ versus $H_d : (C, AF) \sim (1, 0, 0)$, in a subdivided population.

C	AF	PI
$A_i A_i$	$A_i A_i$	$\frac{1+2\theta}{3\theta + (1-\theta)p_i}$
$A_i A_i$	$A_i A_j$	$\frac{1+2\theta}{2[2\theta + (1-\theta)p_i]}$
$A_i A_j$	$A_i A_i$	$\frac{1+2\theta}{2[2\theta + (1-\theta)p_i]}$
$A_i A_j$	$A_i A_j$	$\frac{(1+2\theta)[2\theta + (1-\theta)(p_i + p_j)]}{4[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}$
$A_i A_j$	$A_i A_k$	$\frac{1+2\theta}{4[\theta + (1-\theta)p_i]}$

Table 5.7 The likelihood ratios about two competing hypotheses $H_p : AF$ is the true father of the child versus $H_d : AF$ is a paternal relative of the child, i.e. $H_p : (C, AF) \sim (0, 1, 0)$ versus $H_d : (C, AF) \sim (k_0, 2k_1, 0)$, in a subdivided population, from Fung *et al.* (2003a). (Reproduced by permission of Elsevier.)

C	AF	Likelihood ratio
$A_i A_i$	$A_i A_i$	$\frac{1 + 2\theta}{k_0[3\theta + (1 - \theta)p_i] + 2k_1(1 + 2\theta)}$
$A_i A_i$	$A_i A_j$	$\frac{1 + 2\theta}{2k_0[2\theta + (1 - \theta)p_i] + 2k_1(1 + 2\theta)}$
$A_i A_j$	$A_i A_i$	$\frac{1 + 2\theta}{2k_0[2\theta + (1 - \theta)p_i] + 2k_1(1 + 2\theta)}$
$A_i A_j$	$A_i A_j$	$\frac{(1 + 2\theta)[2\theta + (1 - \theta)(p_i + p_j)]}{4k_0[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j] + 2k_1(1 + 2\theta)[2\theta + (1 - \theta)(p_i + p_j)]}$
$A_i A_j$	$A_i A_k$	$\frac{1 + 2\theta}{4k_0[\theta + (1 - \theta)p_i] + 2k_1(1 + 2\theta)}$

case, where the two competing hypotheses are described by

$$H_p: (C, AF) \sim (0, 1, 0);$$

$$H_d: (C, AF) \sim (k_0, 2k_1, 0).$$

Under H_p , AF is the true father of the child, while under the alternative proposition H_d , the defendant argued that AF is only a paternal relative (such as uncle, say) of the child. The likelihood ratios for all possible genotype combinations of C and AF are listed in Table 5.7, which are derived directly from Table 5.5.

When the alleged father is not available but his relative Z is, we can type the relative instead. The hypotheses of interest are

$$H_p: \text{a relative of } Z \text{ is the true father of the child } Y;$$

$$H_d: \text{the true father is a random unrelated man.}$$

That is,

$$H_p: (Y, Z) \sim (k_0, 2k_1, 0);$$

$$H_d: (Y, Z) \sim (1, 0, 0).$$

If the likelihood ratio for the hypotheses is denoted by LR , then the following simple relationship can be obtained:

$$LR = k_0 + (1 - k_0)PI = (1 - 2k_1) + 2k_1PI, \quad (5.8)$$

which can be verified throughout all of the seven cases listed in Table 5.5. The PI 's are listed in Table 5.6.

Using the notation δ_0 introduced in Evett and Weir (1998), Equation (5.8) can be expressed equivalently as

$$LR = \delta_0 + (1 - \delta_0)PI. \quad (5.9)$$

In fact, δ_0 is defined as the probability that no alleles in the two individuals Y and Z are ibd, which is obviously k_0 . Ayres (2000) reported Equation (5.9) when δ_0 takes the specific values of half siblings and first cousin relationships.

A measure of relatedness θ_{AT} (Evetts and Weir 1998) for individuals AF and the true father TF of the child is defined as the probability that two alleles, one taken at random from each of AF and TF , are ibd, which is just the kinship coefficient F between AF and TF defined in Equation (3.21). Since each of the two alleles of TF is equally likely of being transmitted to the child, θ_{AT} and k_1 are equivalent. So, Equation (5.8) can also be expressed as

$$LR = (1 - 2\theta_{AT}) + 2\theta_{AT} PI. \quad (5.10)$$

Equation (5.10) was reported in the special case that θ is taken as zero (Evetts and Weir 1998), which is applicable in the population when the Hardy–Weinberg law holds.

5.4 Examples with software

The computer software EasyDNA_2Persons discussed earlier in Section 5.2 can also be used for kinship testing of any two persons in a subdivided population. The steps in running the software are exactly the same as those listed in Section 5.2 except for the addition of a step 1a after step 1:

- 1a Click the *theta* button and input the appropriate θ .

The program gives the likelihood ratios under Hardy–Weinberg equilibrium ($\theta = 0$) and under a subdivided population having the chosen θ value. Two real case examples reported in Fung *et al.* (2003a) are considered.

The first case of disputed paternity testing was provided by the Hong Kong Government Laboratory, where 12 STR loci (D3S1358, vWA, FGA, D5S818, D13S317, D7S820, D8S1179, D21S11, D18S51, THO1, TPOX, CSF1PO) were typed for a child and an alleged father [see Table 5.8 for information on the genotypes and, for the allele frequencies, one can refer to Wong *et al.* (2001)]. The following four hypotheses are proposed to describe the relationship between the alleged father and the child:

H_p : the alleged father is the true father of the child;

H_{d1} : the alleged father is unrelated to the child;

H_{d2} : the alleged father is the uncle of the child;

H_{d3} : the alleged father and the child are full siblings.

The example is analyzed using the EasyDNA_2Persons software. The overall likelihood ratios about H_p versus H_{di} , $i = 1, 2, 3$ for various values of θ are also obtained and they are shown in Table 5.9. In testing H_p versus H_{d1} , the PI when $\theta = 0$ is equal to 428, which is not too large. The PI decreases when the population structure is taken into account. When $\theta = 0.03$, the PI drops by about 70% to 136 and the genetic evidence may not be strong enough for paternity.

If we are testing H_p against other relationships such as uncle and nephew (H_{d2}), the genetic evidence would become much weaker. The PI is only 9.7 when $\theta = 0$, and is even smaller

Table 5.8 The genotypes of the alleged father and the child of the two disputed paternity testing cases in Hong Kong and Spain, from Fung *et al.* (2003a). (Reproduced by permission of Elsevier.)

Locus	Hong Kong		Spain	
	Alleged father	Child	Alleged father	Child
D3S1358	15/16	16/17	16/18	16/18
vWA	14/18	14/18	16/18	18/18
FGA	22/24	20/24	22/24	21/22
THO1	7/9	6/7	9/9.3	7/9.3
TPOX	11/11	8/11	8/9	9/11
CSF1PO	10/12	12/12	9/12	9/11
D5S818	10/11	11/12	11/11	11/13
D13S317	9/10	10/10	12/13	8/12
D7S820	8/11	8/8	11/11	11/11
D8S1179	11/14	14/16	10/13	10/13
D21S11	29/32.2	30/32.2	30/30	30/31
D18S51	13/14	13/14	14/14	14/18
D16S539	—	—	9/11	9/12

when the population subdivision is taken into account. Similar findings are observed when we are testing H_p versus H_{d3} .

The second case of disputed paternity testing comes from a Spanish population. The same set of 12 STR loci with an additional locus D16S539 were typed for a child and an alleged father [also see Table 5.8 for the genotypes and, for the allele frequencies, one can refer to Gusmão *et al.* (2000)]. The same sets of hypotheses are chosen.

In testing H_p versus H_{d1} , unlike the case in Hong Kong, the overall PI when $\theta = 0$ is equal to 135 689 (see Table 5.9) which is very large, giving very strong evidence for paternity. When we increase the value of θ from 0.01, 0.02 to 0.03, the PI becomes 30, 12 and 5.9% of the PI when $\theta = 0$. The value drops substantially with the increase in θ . If we investigate the individual PI value (not shown) at each locus, we notice that for all loci, except CSF1PO,

Table 5.9 Likelihood ratios with different θ values for the two disputed paternity testing cases in Hong Kong and Spain.

Hypotheses ^a	$\theta = 0$	0.01	0.02	0.03
(Hong Kong)				
H_p versus H_{d1}	428	280	192	136
H_p versus H_{d2}	9.7	8.6	7.6	6.8
H_p versus H_{d3}	12.2	11.9	11.6	11.4
(Spain)				
H_p versus H_{d1}	135 689	41 148	16 766	7 989
H_p versus H_{d2}	45.5	37.5	31.2	26.2
H_p versus H_{d3}	12.4	12.4	12.4	12.4

^a H_p : father and son; H_{d1} : unrelated persons; H_{d2} : uncle and nephew; H_{d3} : full siblings

the PI drops by a few percent to about 30% when θ is increased from 0 to 0.03. The PI value at CSF1PO, however, has a large drop from 20.8 to 11.7, 8.2 and 6.4 for $\theta = 0.01, 0.02$ and 0.03 , respectively. If we look at the genotypes of the alleged father and the child, we find that they share one common allele (9) at CSF1PO, which is a very rare allele in the Spanish population (with frequency 0.012). The value of θ has a crucial effect on PI for the case with a rare allele.

Table 5.9 also shows the PI 's for the other two sets of hypotheses. From both paternity cases in Hong Kong and Spain, we notice that the paternity indices for testing H_p (father and son) versus H_{d3} (full siblings) have values of about 10. This phenomenon is not unusual, and it may have implications for some paternity cases such as those found in immigration (Fung *et al.* 2002, see also Section 4.9.2).

In the majority of paternity cases, PI values are so high that even if there is strong population structure, the final consequences will generally be of little importance. But, in deficiency paternity testing and many immigration cases, the effect of population subdivision may be of importance and mistakes could be made if this effect is not properly taken into account in the calculations.

5.5 Three persons situation: HWE

The testing for a biological relationship between two persons in a population with Hardy–Weinberg equilibrium was studied in detail by Li and Sacks (1954). Ayres (2000) and Fung *et al.* (2003a) have successfully generalized Li and Sack's results to a subdivided population. In the following, we consider an extension of the results to a three persons situation.

Let X, Z denote the maternal and paternal relatives of Y , respectively, and suppose that X and Z are unrelated. The relatedness coefficients between X and Y , and between Y and Z are denoted as $(k_0^{XY}, 2k_1^{XY}, 0)$ and $(k_0^{YZ}, 2k_1^{YZ}, 0)$, respectively. The population is assumed to be in Hardy–Weinberg equilibrium. In order to determine the biological relationship among the three individuals X, Y , and Z , we list in Table 5.10 all the possible joint genotype probabilities $P(X, Y, Z)$.

To see the derivation of the joint genotype probabilities given in Table 5.10, we consider $Y = A_i A_i$, $X = A_i A_j$ and $Z = A_k A_l$, $j \neq i$, $k \neq i$, $l \neq i$. It is concluded from the genotypes of X, Y and Z that X and Y may share i th allele A_i , and Y and Z share no i th alleles. The joint genotype probability can then be evaluated as

$$\begin{aligned}
 &P(X = A_i A_j, Y = A_i A_i, Z = A_k l) \\
 &= P(X = A_i A_j, Y = A_i A_i) P(Z = A_k l) \times k_0^{YZ} \\
 &= \sum_{t=0}^2 P(X = A_i A_j, Y = A_i A_i | X \text{ and } Y \text{ share } t \text{ ibd alleles}) \\
 &\quad \times P(X \text{ and } Y \text{ share } t \text{ ibd alleles}) \times P(Z = A_k l) \times k_0^{YZ} \\
 &= [P(X = A_i A_j) P(Y = A_i A_i) k_0^{XY} + P(A_i, A_i, A_j) \times 2k_1^{XY}] \times P(Z = A_k A_l) \times k_0^{YZ} \\
 &= k_0^{XY} k_0^{YZ} p_i^2 P(X = A_i A_j) P(Z = A_k A_l) + 2k_1^{XY} k_0^{YZ} p_i^2 p_j P(Z = A_k A_l) \\
 &= k_0^{XY} k_0^{YZ} p_i^2 P(X) P(Z) + 2k_1^{XY} k_0^{YZ} p_i^2 p_j P(Z), \tag{5.11}
 \end{aligned}$$

where $P(X)$ and $P(Z)$ are abbreviated forms of $P(X = A_i A_j)$ and $P(Z = A_k A_l)$, respectively.

Table 5.10 The joint genotype probabilities $P(X, Y, Z)$, for all possible combinations of X, Y and Z (regardless of order X and Z), where X and Z are the maternal and paternal relatives of Y , respectively; X and Z are unrelated; $(k_0^{XY}, 2k_1^{XY}, 0)$ are the relatedness coefficients of X and Y ; and $(k_0^{YZ}, 2k_1^{YZ}, 0)$ are the relatedness coefficients of Y and Z , from Fung *et al.* (2006). (Reproduced by permission of Elsevier.)

Y	X	Z	$P(X, Y, Z)$
$A_i A_i$	$A_i A_j$	$A_i A_k$	$k_0^{XY} k_0^{YZ} p_i^2 P(X) P(Z) + 2k_0^{XY} k_1^{YZ} p_i^2 p_k P(X)$ $+ 2k_1^{XY} k_0^{YZ} p_i^2 p_j P(Z) + 4k_1^{XY} k_1^{YZ} p_i^2 p_j p_k$
	$A_i A_j$	$A_k A_l$	$k_0^{XY} k_0^{YZ} p_i^2 P(X) P(Z) + 2k_1^{XY} k_0^{YZ} p_i^2 p_j P(Z)$
		$(k, l \neq i)$	
	$A_j A_k$	$A_l A_m$	$k_0^{XY} k_0^{YZ} p_i^2 P(X) P(Z)$
	$(j, k \neq i)$	$(l, m \neq i)$	
$A_i A_j$	$A_i A_j$	$A_i A_j$	$8k_0^{XY} k_0^{YZ} p_i^3 p_j^3 + 4k_0^{XY} k_1^{YZ} (p_i + p_j) p_i^2 p_j^2$ $+ 4k_1^{XY} k_0^{YZ} (p_i + p_j) p_i^2 p_j^2 + 8k_1^{XY} k_1^{YZ} p_i^2 p_j^2$
$(j \neq i)$	$(j \neq i)$	$(j \neq i)$	
		$A_i A_k$	$4k_0^{XY} k_0^{YZ} p_i^2 p_j^2 P(Z) + 4k_0^{XY} k_1^{YZ} p_i^2 p_j^2 p_k$ $+ 2k_1^{XY} k_0^{YZ} (p_i + p_j) p_i p_j P(Z) + 4k_1^{XY} k_1^{YZ} p_i^2 p_j p_k$
		$(k \neq j)$	
		$A_k A_l$	$4k_0^{XY} k_0^{YZ} p_i^2 p_j^2 P(Z)$
		$(k, l \neq i, j)$	$+ 2k_1^{XY} k_0^{YZ} (p_i + p_j) p_i p_j P(Z)$
	$A_i A_k$	$A_i A_l$	$2k_0^{XY} k_0^{YZ} p_i p_j P(X) P(Z) + 2k_0^{XY} k_1^{YZ} p_i p_j p_l P(X)$ $+ 2k_1^{XY} k_0^{YZ} p_i p_j p_k P(Z)$
	$(k \neq j)$	$(l \neq j)$	
		$A_j A_l$	$2k_0^{XY} k_0^{YZ} p_i p_j P(X) P(Z) + 2k_0^{XY} k_1^{YZ} p_i p_j p_l P(X)$ $+ 2k_1^{XY} k_0^{YZ} p_i p_j p_k P(Z) + 4k_1^{XY} k_1^{YZ} p_i p_j p_k p_l$
		$(l \neq i)$	
		$A_l A_m$	$2k_0^{XY} k_0^{YZ} p_i p_j P(X) P(Z) + 2k_1^{XY} k_0^{YZ} p_i p_j p_k P(Z)$
		$(l, m \neq i, j)$	
	$A_k A_l$	$A_m A_n$	$2k_0^{XY} k_0^{YZ} p_i p_j P(X) P(Z)$
	$(k, l \neq i, j)$	$(m, n \neq i, j)$	

The above derivation considers the situation that $j \neq i$. When $j = i$, i.e. $X = A_i A_i$ and $Y = A_i A_i$, it looks as though X and Y may share two ibd A_i alleles. This is, however, impossible, since X and Y are only maternally related and so they cannot share two ibd alleles, i.e. $k_2^{XY} = 0$. Hence, the same formula in Equation (5.11) results when $j = i$. Thus, we have obtained the joint genotype probability Equation (5.11) for $Y = A_i A_i$, $X = A_i A_j$ and $Z = A_k A_l$, $k, l \neq i$, which is shown in the second row of Table 5.10.

The other joint genotype probabilities in Table 5.10 can be shown in a similar way [see also Fung *et al.* (2006)]. The details are omitted for brevity. The expressions in Table 5.10 can be used to examine the biological relationship among any three persons X , Y , and Z where X and Z are unrelated. For example, we can use one maternal relative and one paternal relative to identify a missing person. We can also examine the biological relationship between Y and Z when the biological relationship between X and Y are known without error such as in the paternity testing case.

For two competing hypotheses H_p and H_d , the likelihood ratio is

$$LR = \frac{P(\text{evidence}|H_p)}{P(\text{evidence}|H_d)} = \frac{P(X, Y, Z|H_p)}{P(X, Y, Z|H_d)}. \quad (5.12)$$

It is a ratio of two like genotype probabilities, one evaluated under H_p and the other under H_d . These probabilities can be obtained from Table 5.10, with values of $(k_0^{XY}, 2k_1^{XY}, k_0^{YZ}, 2k_1^{YZ})$ determined under H_p and H_d . To illustrate this, we consider a paternity testing problem with genotypes of three persons provided in Table 5.11. Person Y is always regarded as the child in the following consideration:

- (i) X is the mother of Y and Z is the alleged father with propositions

$$\begin{aligned} H_p : Z \text{ is the true father of the child } Y; \\ H_d : \text{the true father is a random unrelated man.} \end{aligned} \quad (5.13)$$

In this case, the relatedness coefficients for X and Y are $k_0^{XY} = 0$ and $2k_1^{XY} = 1$, and for Y and Z are $k_0^{YZ} = 0$ and $2k_1^{YZ} = 1$. Plugging them into the appropriate formulas in Table 5.10 can provide the likelihood ratio we want. In fact, this is just a standard trio problem. The overall likelihood ratio for genotypes in Table 5.11 is obtained as 37.36. The details are omitted for brevity.

- (ii) The mother is unavailable but her brother is. Then X is the uncle of the child Y and Z is the alleged father. The propositions are the same as those listed in (5.13). The relatedness coefficients are

$$\begin{aligned} k_0^{XY} = 2k_1^{XY} = 0.5, \quad k_0^{YZ} = 0 \quad \text{and} \quad 2k_1^{YZ} = 1, \quad \text{under } H_p; \\ k_0^{XY} = 2k_1^{XY} = 0.5, \quad k_0^{YZ} = 1 \quad \text{and} \quad 2k_1^{YZ} = 0, \quad \text{under } H_d. \end{aligned}$$

Table 5.11 Genotype data of three persons Y , X and Z at loci D3S1358, wWA and FGA.

Locus	Y	X	Z
D3S1358	15/17	15/16	17/18
vWA	18/18	18/19	18/18
FGA	20/21	20/21	19/20

Suppose we write the likelihood ratio in Equation (5.12) as $LR = Num/Den$. Based on the formulas in Table 5.10 and the genotypes in Table 5.11, we have

$$\begin{aligned}
 \text{D3S1358 : } Num &= 2k_0^{XY}k_0^{YZ}p_i p_j P(X)P(Z) + 2k_0^{XY}k_1^{YZ}p_i p_j p_l P(X) \\
 &\quad + 2k_1^{XY}k_0^{YZ}p_i p_j p_k P(Z) + 4k_1^{XY}k_1^{YZ}p_i p_j p_k p_l \\
 &= 2 \times 0.5 \times 0 + 2 \times 0.5 \times 0.5 \times p_{15}p_{17}p_{18} \times 2p_{15}p_{16} \\
 &\quad + 0.5 \times 0 + 0.5 \times p_{15}p_{17}p_{16}p_{18} \\
 &= 2 \times 0.5 \times 0.5 \times 0.331 \times 0.239 \times 0.056 \times 2 \times 0.331 \\
 &\quad \times 0.326 + 0.5 \times 0.331 \times 0.239 \times 0.326 \times 0.056 \\
 &= 1.20 \times 10^{-3}, \\
 Den &= 2 \times 0.5 \times 1 \times 0.331 \times 0.239 \times 2 \times 0.331 \times 0.326 \\
 &\quad \times 2 \times 0.239 \times 0.056 + 0.5 \times 0 + 0.5 \times 1 \times 0.331 \\
 &\quad \times 0.239 \times 0.326 \times 2 \times 0.239 \times 0.056 + 2 \times 0.5 \times 0 \\
 &= 8.02 \times 10^{-4}, \\
 LR &= 1.20 \times 10^{-3} / (8.02 \times 10^{-4}) = 1.50,
 \end{aligned}$$

and the likelihood ratios for vWA and FGA can be obtained similarly. They are

$$\text{vWA : } LR = 6.25, \quad \text{and} \quad \text{FGA : } LR = 3.19,$$

giving an overall likelihood ratio $LR = 29.9$.

- (iii) Both the mother and alleged father are unavailable, but the brother of the mother and the father of the alleged father are. Then X is the maternal uncle of the child Y , and Z is the paternal grandfather of the child under H_p . Thus, the propositions become

$$\begin{aligned}
 H_p &: Z \text{ is the paternal grandfather of the child } Y; \\
 H_d &: \text{the true father is a random unrelated man.}
 \end{aligned} \tag{5.14}$$

The relatedness coefficients are

$$\begin{aligned}
 k_0^{XY} &= 2k_1^{XY} = 0.5, \quad k_0^{YZ} = 2k_1^{YZ} = 0.5, \quad \text{under } H_p; \\
 k_0^{XY} &= 2k_1^{XY} = 0.5, \quad k_0^{YZ} = 1 \text{ and } 2k_1^{YZ} = 0, \quad \text{under } H_d.
 \end{aligned}$$

Based on the formulas in Table 5.10 and the genotypes in Table 5.11, we have

$$\begin{aligned}
 \text{D3S1358 : } Num &= 2k_0^{XY}k_0^{YZ}p_i p_j P(X)P(Z) + 2k_0^{XY}k_1^{YZ}p_i p_j p_l P(X) \\
 &\quad + 2k_1^{XY}k_0^{YZ}p_i p_j p_k P(Z) + 4k_1^{XY}k_1^{YZ}p_i p_j p_k p_l \\
 &= 2 \times 0.5 \times 0.5 \times 0.331 \times 0.239 \times 2 \times 0.331 \\
 &\quad \times 0.326 \times 2 \times 0.239 \times 0.056 + 0.5 \times 0.5 \\
 &\quad \times 0.331 \times 0.239 \times 0.056 \times 2 \times 0.331 \times 0.326 \\
 &\quad + 0.5 \times 0.5 \times 0.331 \times 0.239 \times 0.326
 \end{aligned}$$

$$\begin{aligned}
& \times 2 \times 0.239 \times 0.056 + 0.5 \times 0.5 \times 0.331 \\
& \times 0.239 \times 0.326 \times 0.056 \\
& = 1.001 \times 10^{-3}, \\
Den & = 8.02 \times 10^{-4}, \text{ as obtained previously in (ii),} \\
LR & = 1.001 \times 10^{-3} / (8.02 \times 10^{-4}) = 1.25,
\end{aligned}$$

and the likelihood ratios at vWA and FGA can be obtained similarly. They are

$$vWA : LR = 3.62, \quad \text{and} \quad FGA : LR = 2.09,$$

giving an overall likelihood ratio $LR = 9.46$.

As we can see from previous discussions, the overall likelihood ratio is the highest in case (i) of the standard trio problem, becomes smaller in case (ii) in which the mother of the child is not available for typing but her brother is, and drops again in case (iii), in which, additionally, the alleged father is unavailable but his father is. In other words, the likelihood ratio becomes smaller and smaller when the biological relationship between the child and his/her maternal and paternal relatives who provide genotyping information becomes looser and looser. This phenomenon is generally true in kinship testing.

5.6 Computer software and example

Although we have extended Li and Sacks' (1954) joint genotype probability for two persons to the case of three persons X , Y and Z , where X and Z are biologically unrelated, the formulas given in Table 5.10 are, however, rather lengthy. We have developed a computer software named EasyDNA_3Persons to deal with the associated kinship problems.

Steps in running the EasyDNA 3Persons software

- 1 Click the *Load frequency file* button after loading the EasyDNA program, then select the appropriate file
- 2 Choose the allele pairs at the locus for Y , X and Z
- 3a Choose the appropriate maternal relation between Y and X under H_p [which is, for the example in case (iii) above, *Nephew-uncle*]
- 3b Choose the appropriate paternal relation between Y and Z under H_p [which is, for the example in case (iii) above, *Child-grandparent*]
- 3c Choose the appropriate maternal relation between Y and X under H_d [which is, for the example in case (iii) above, *Nephew-uncle*]
- 3d Choose the appropriate paternal relation between Y and Z under H_d [which is, for the example in case (iii) above, *unrelated*]
- 4 Click the *Calculate* button
- 5 Repeat steps 2 and 4 for each of the remaining loci; step 3 (3a–3d) is blocked, since it is no longer needed for the remaining loci.

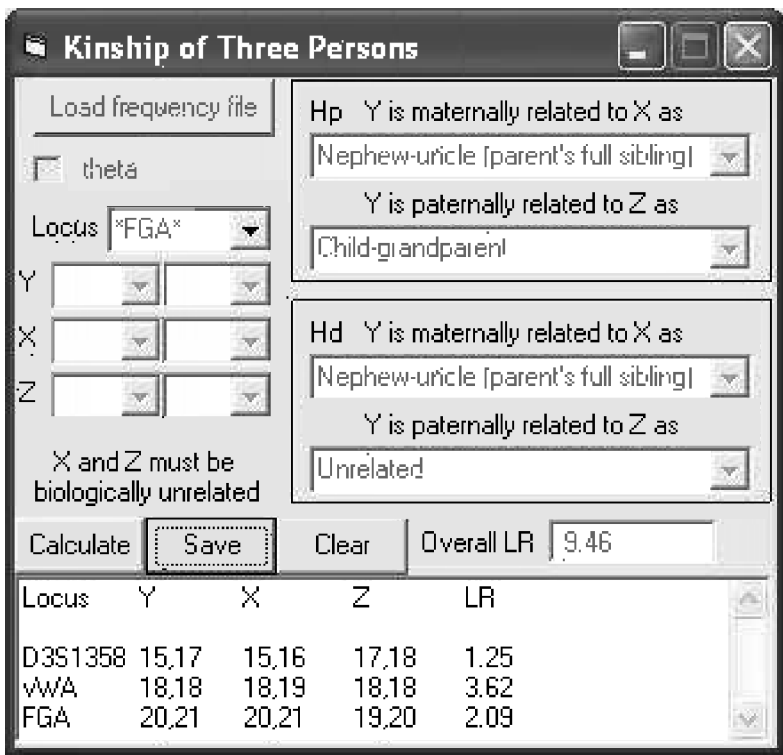


Figure 5.3 Captured screen for running the EasyDNA_3Persons software for analyzing the example given in case (iii) of Section 5.5.

The procedure steps are straightforward and easy to follow.

We use the software to deal with the example in case (iii) above. Figure 5.3 gives the screen capture in the running of the program. Although we have shown earlier that the numerical calculations are rather tedious, we can see from the figure that our EasyDNA_3Persons can handle the problem easily. The likelihood ratios are obtained as 1.25, 3.62 and 2.09 for individual loci, giving an overall ratio of 9.46.

5.7 Three persons situation: subdivided populations

5.7.1 Standard trio

Consider the trio problem in which the DNA profiles of the child, mother and alleged father are typed. Suppose that the mother, alleged father and biological father of the child are assumed to be unrelated to one another. The usual hypotheses are given as

$$\begin{aligned} H_p &: \text{the alleged father is the true father of the child;} \\ H_d &: \text{the true father is a random unrelated man.} \end{aligned} \tag{5.15}$$

Let M , C and AF be the genotypes of the mother, her child and the alleged father, respectively. The paternity index can be expressed as (Section 4.1.1)

$$PI = \frac{P(C|M, AF, H_p)}{P(C|M, AF, H_d)}. \quad (5.16)$$

Consider a particular locus that $C = A_i A_j$, $M = A_i A_k$ and $AF = A_j A_l$. In this situation, the numerator of the paternity index is $(1/2) \times (1/2)$. However, the denominator is not $(1/2)p_j$ because the population is not in Hardy–Weinberg equilibrium. It is in fact equal to

$$P(C = A_i A_j | M = A_i A_k, AF = A_j A_l, H_d) = (1/2)P(A_j | A_i, A_k, A_j, A_l, H_d),$$

since the mother has $1/2$ chance of passing A_i to the child. The conditional probability of observing the other allele A_j for the child is obtained according to Equation (3.17). The denominator then becomes

$$(1/2)[\theta + (1 - \theta)p_j]/[1 + (4 - 1)\theta].$$

Thus, the paternity index is

$$\begin{aligned} PI &= \frac{(1/4)}{(1/2)[\theta + (1 - \theta)p_j]/(1 + 3\theta)} \\ &= \frac{(1 + 3\theta)}{2[\theta + (1 - \theta)p_j]}. \end{aligned}$$

The paternity indices for other combinations of genotypes of C , M and AF can be derived similarly, and they are all shown in Table 5.12 [see also Evett and Weir (1998)].

5.7.2 A relative of the alleged father is the true father

In the situation in which the alleged father gives an alternative explanation that his relative is the true father of the child, the defense proposition becomes

H_d : a relative of the alleged father is the true father of the child.

The prosecution proposition H_p remains unchanged, as in (5.15). Suppose that the genotypes of the child, mother and alleged father are obtained as $C = A_i A_j$, $M = A_i A_k$ and $AF = A_j A_l$, respectively. The numerator of the likelihood ratio in Equation (5.16) is $(1/2) \times (1/2) = 1/4$. For the denominator, it is concluded from the genotypes of the child and the mother that $C_M = A_i$ and $C_P = A_j$. So

$$\begin{aligned} P(C = A_i A_j | M = A_i A_k, AF = A_j A_l, H_d) \\ &= P(C_M = A_i, C_P = A_j | M = A_i A_k, AF = A_j A_l, H_d) \\ &= (1/2)P(C_P = A_j | M = A_i A_k, AF = A_j A_l, H_d). \end{aligned}$$

Let $(k_0, 2k_1, k_2)$ be the relatedness coefficients between C and AF . Consider the probability that $C_P = A_j$ is not ibd with the A_j allele of AF is k_0 and the probability that $C_P = A_j$ is ibd with the A_j allele of AF is k_1 , so

$$\begin{aligned} P(C_P = A_j | M = A_i A_k, AF = A_j A_l, H_d) \\ &= k_0 P(A_j | A_i, A_j, A_k, A_l) + k_1 \\ &= k_0 \frac{\theta + (1 - \theta)p_j}{1 + 3\theta} + k_1. \end{aligned}$$

Table 5.12 Paternity index (PI) for a standard trio case in a subdivided population.

C	M	AF	PI
$A_i A_i$	$A_i A_i$	$A_i A_i$	$\frac{1 + 3\theta}{4\theta + (1 - \theta)p_i}$
		$A_i A_j$	$\frac{1 + 3\theta}{2[3\theta + (1 - \theta)p_i]}$
		$A_i A_k$	$\frac{1 + 3\theta}{2[2\theta + (1 - \theta)p_i]}$
	$A_i A_j$	$A_i A_i$	$\frac{1 + 3\theta}{3\theta + (1 - \theta)p_i}$
		$A_i A_j$	$\frac{1 + 3\theta}{2[2\theta + (1 - \theta)p_i]}$
		$A_i A_k$	$\frac{1 + 3\theta}{2[2\theta + (1 - \theta)p_i]}$
	$A_i A_k$	$A_i A_i$	$\frac{1 + 3\theta}{2[\theta + (1 - \theta)p_i]}$
		$A_i A_j$	$\frac{1 + 3\theta}{2\theta + (1 - \theta)p_j}$
		$A_i A_k$	$\frac{1 + 3\theta}{2[\theta + (1 - \theta)p_j]}$
$A_i A_j$	$A_i A_i$	$A_i A_j$	$\frac{1 + 3\theta}{4\theta + (1 - \theta)(p_i + p_j)}$
		$A_i A_k$	$\frac{1 + 3\theta}{4\theta + (1 - \theta)(p_i + p_j)}$
		$A_i A_l$	$\frac{1 + 3\theta}{2[3\theta + (1 - \theta)(p_i + p_j)]}$
	$A_i A_j$	$A_i A_i$	$\frac{1 + 3\theta}{2[\theta + (1 - \theta)p_j]}$
		$A_i A_j$	$\frac{1 + 3\theta}{2\theta + (1 - \theta)p_j}$
		$A_i A_k$	$\frac{1 + 3\theta}{2[\theta + (1 - \theta)p_j]}$
	$A_i A_k$	$A_i A_j$	$\frac{1 + 3\theta}{2k_1(1 + 3\theta) + 2k_0[\theta + (1 - \theta)p_j]}$
		$A_i A_k$	$\frac{1 + 3\theta}{2k_1(1 + 3\theta) + 2k_0[\theta + (1 - \theta)p_j]}$
		$A_i A_l$	$\frac{1 + 3\theta}{2k_1(1 + 3\theta) + 2k_0[\theta + (1 - \theta)p_j]}$

If follows that the likelihood ratio is

$$\begin{aligned}
 LR &= \frac{1/4}{(1/2)[k_1 + k_0 \frac{\theta + (1 - \theta)p_j}{1 + 3\theta}]} \\
 &= \frac{1 + 3\theta}{2k_1(1 + 3\theta) + 2k_0[\theta + (1 - \theta)p_j]}
 \end{aligned}$$

$$= \frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[\theta + (1 - \theta)p_j]},$$

where F is the kinship coefficient between the alleged father and his relative who is the true father of the child under H_d (see Section 4.3). This likelihood ratio is reported in the last row of Table 5.13.

The likelihood ratios for all other genotype combinations of C , M and AF can be derived similarly. These ratios are all shown in Table 5.13. In particular, we also show the likelihood ratios for the defense proposition H_d^* : a brother of the alleged father is the true father of the child, in which the kinship coefficient between the alleged father and the true father is $F = 1/4$. The ratios for this particular H_d^* proposition are presented in the last column of Table 5.13, and have also been reported by Buckleton *et al.* (2005).

5.7.3 Alleged father unavailable but his relative is

When the alleged father is not available and his relative R is tested instead, then the hypotheses of interest are given as

H_p : a relative of R is the true father of the child;

H_d : the true father is a random unrelated man.

Based on a similar argument as presented in Section 4.4, in which Hardy–Weinberg equilibrium holds, we can derive the avuncular index for the above hypotheses as

$$AI = \frac{PI}{LR},$$

where PI and LR are provided in Tables 5.12 and 5.13, respectively, and the genotypes of R are listed as those shown in the AF columns.

Morris *et al.* (1988) noticed the following relationship under the standard trio case with Hardy–Weinberg equilibrium:

$$AI = (1 - 2F) + 2F \times PI.$$

After some simple derivations, this relationship can be shown to hold true in the current situation in which the mother, alleged father, R and true father belong to the same subdivided population. In addition to the findings we obtained earlier for the no-mother case, this relationship is found to hold true for populations with Hardy–Weinberg equilibrium or with subdivision, and in the with-mother or without-mother case.

5.7.4 Example

We consider the standard trio example given in Table 4.2 in which the mother, alleged father and true father come from a subdivided population. The value of θ is taken to be 0.01. The hypotheses of interest are

H_p : the alleged father is the true father of the child;

H_d : the true father is a random unrelated man.

(5.17)

Based on the formulas given in Table 5.12, the paternity indices are obtained as

$$\text{D3S1358 : } PI = \frac{1 + 3\theta}{2[\theta + (1 - \theta)p_{17}]} = \frac{1 + 3 \times 0.01}{2 \times [0.01 + (1 - 0.01) \times 0.239]} = 2.088,$$

Table 5.13 Likelihood ratio for a trio case in a subdivided population with H_p : the alleged father is the true father of the child versus H_d : a relative of the alleged father is the true father; F is the kinship coefficient for the alleged father and his relative who is the true father under H_d , with $F = 1/4$ being the coefficient for brothers.

C	M	AF	LR	$LR(F = 1/4)$
$A_i A_i$	$A_i A_i$	$A_i A_i$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + (1 - 2F)[4\theta + (1 - \theta)p_i]}$	$\frac{2(1 + 3\theta)}{1 + 7\theta + (1 - \theta)p_i}$
		$A_i A_j$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[3\theta + (1 - \theta)p_i]}$	$\frac{2(1 + 3\theta)}{1 + 9\theta + 2(1 - \theta)p_i}$
		$A_i A_j$	$A_i A_i$	$\frac{1 + 6\theta + (1 - \theta)p_i}{2(1 + 3\theta)}$
			$A_i A_j$	$\frac{1 + 7\theta + 2(1 - \theta)p_i}{2(1 + 3\theta)}$
			$A_i A_k$	$\frac{1 + 7\theta + 2(1 - \theta)p_i}{2(1 + 3\theta)}$
		$A_i A_j$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[\theta + (1 - \theta)p_j]}$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$
	$A_i A_j$	$A_i A_i$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + (1 - 2F)[4\theta + (1 - \theta)p_0^*]}$	$\frac{1 + 7\theta + (1 - \theta)p_0}{2(1 + 3\theta)}$
		$A_i A_j$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + (1 - 2F)[4\theta + (1 - \theta)p_0]}$	$\frac{1 + 7\theta + (1 - \theta)p_0}{2(1 + 3\theta)}$
		$A_i A_k$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[3\theta + (1 - \theta)p_0]}$	$\frac{1 + 9\theta + 2(1 - \theta)p_0}{2(1 + 3\theta)}$
		$A_i A_k$	$A_i A_j$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$
			$A_j A_j$	$\frac{1 + 5\theta + (1 - \theta)p_j}{2(1 + 3\theta)}$
			$A_j A_k$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$
	$A_i A_k$	$A_i A_j$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[\theta + (1 - \theta)p_j]}$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$
		$A_j A_j$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + (1 - 2F)[2\theta + (1 - \theta)p_j]}$	$\frac{1 + 5\theta + (1 - \theta)p_j}{2(1 + 3\theta)}$
		$A_j A_k$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[\theta + (1 - \theta)p_j]}$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$
		$A_j A_l$	$\frac{1 + 3\theta}{2F(1 + 3\theta) + 2(1 - 2F)[\theta + (1 - \theta)p_j]}$	$\frac{1 + 5\theta + 2(1 - \theta)p_j}{2(1 + 3\theta)}$

* $p_0 = p_i + p_j$

$$\text{FGA : } PI = \frac{1 + 3\theta}{3\theta + (1 - \theta)p_{18}} = \frac{1 + 3 \times 0.01}{3 \times 0.01 + (1 - 0.01) \times 0.160} = 5.467,$$

$$\begin{aligned} \text{vWA : } PI &= \frac{1 + 3\theta}{2[3\theta + (1 - \theta)(p_{20} + p_{21})]} \\ &= \frac{1 + 3 \times 0.01}{2 \times [3 \times 0.01 + (1 - 0.01) \times (0.044 + 0.131)]} \\ &= 2.534. \end{aligned}$$

The overall paternity index is 28.93.

Suppose that the defendant provides an alternative explanation that his brother is the true father of the child. Then, the defense proposition becomes

$$H_d^* : \text{the brother of the alleged father is the true father of the child.} \quad (5.18)$$

In this situation, the child and the alleged father are nephew–uncle related under H_d , i.e. $(k_0, 2k_1, k_2) = (0.5, 0.5, 0)$. Using the formulas given in Table 5.13, we obtain the likelihood ratios

$$\text{D3S1358} : LR = 1.352,$$

$$\text{FGA} : LR = 1.691,$$

$$\text{vWA} : LR = 1.434.$$

The overall likelihood ratio is 3.278, which is (much) smaller than 28.93—the likelihood ratio under the alternative explanation H_d given in (5.17). In other words, to the defendant, the DNA evidence against him is weaker under the alternative explanation H_d^* that his brother is the true father of the child.

5.7.5 General method and computer software

The joint genotype probabilities for the special three person cases under Hardy–Weinberg equilibrium can be categorized into 10 different scenarios and they have been summarized in Table 5.10. However, when there is population subdivision, the situations become much more complicated. To illustrate the way of obtaining the joint genotype probability, we consider $Y = A_i A_i$, $X = A_i A_j$ and $Z = A_k A_k$, where i, j, k are all distinct. Recall that Y and X are taken to be maternally related, Y and Z paternally related, while X and Z are unrelated. It is clear from the genotypes of X, Y and Z that X and Y may share an ibd allele A_i , and Y and Z share no ibd alleles. If X, Y and Z come from the same subdivided population, using the law of total probability, the joint genotype probability can be obtained as

$$\begin{aligned} P(X = A_i A_j, Y = A_i A_i, Z = A_k A_k) \\ &= \sum_{t=0}^2 P(X = A_i A_j, Y = A_i A_i, Z = A_k A_k | Y \text{ and } Z \text{ share } t \text{ ibd alleles}) \\ &\quad \times P(Y \text{ and } Z \text{ share } t \text{ ibd alleles}) \\ &= P(X = A_i A_j, Y = A_i A_i, A_k, A_k) k_0^{YZ}. \end{aligned}$$

Again, using the law of total probability, it becomes

$$[2k_0^{XY} P(A_i, A_j, A_i, A_i, A_k, A_k) + 2k_1^{XY} P(A_i, A_j, A_i, A_k, A_k)] k_0^{YZ}.$$

Apply the conditional probability formula in Equation (3.17) recursively to $P(A_i, A_j, A_i, A_i, A_k, A_k)$ and $P(A_i, A_j, A_i, A_k, A_k)$, and the joint genotype probability can be obtained as

$$\begin{aligned} P(X = A_i A_j, Y = A_i A_i, Z = A_k A_k) &= 2k_0^{XY} k_0^{YZ} \\ &\times \frac{(1 - \theta) p_i [\theta + (1 - \theta) p_i] [2\theta + (1 - \theta) p_i] (1 - \theta) p_j (1 - \theta) p_k [\theta + (1 - \theta) p_k]}{(1 - \theta) \times 1 \times (1 + \theta) (1 + 2\theta) (1 + 3\theta) (1 + 4\theta)} \end{aligned}$$

$$+ 2k_1^{XY}k_0^{YZ} \times \frac{(1 - \theta)p_i[\theta + (1 - \theta)p_i](1 - \theta)p_j(1 - \theta)p_k[\theta + (1 - \theta)p_k]}{(1 - \theta) \times 1 \times (1 + \theta)(1 + 2\theta)(1 + 3\theta)}.$$

The probabilities for other combinations of genotypes of X , Y and Z can be obtained in similar ways. In fact, there are more than 100 such combinations, and it is not feasible to list all these probabilities. Instead, we have developed a computer software which incorporates all these possibilities. The software can be used to test for kinship among three persons X , Y and Z who belong to the same subdivided population, and X and Z are taken to be unrelated.

In fact, the software EasyDNA_3Persons mentioned in Section 5.6 can deal with the problems in subdivided populations. The steps in running the software are exactly the same as those listed in Section 5.6, except for the addition of the following step:

- 1a Click the *theta* button and input the appropriate θ .

We illustrate using the data set given in Table 5.11, in which Y is regarded as the child and X is the half-brother of the mother of the child. The following hypotheses are of interest:

- H_p : the alleged father Z is the true father of Y ;
 - H_{d1} : the true father is a random man;
 - H_{d2} : the brother of the alleged father is the true father.
- (5.19)

H_p versus H_{d1} , and H_p versus H_{d2} with $\theta = 0.01$ are considered respectively. The likelihood ratios are summarized in Table 5.14. We notice that, as usual, the likelihood ratio decreases when θ increases. This is true for all loci except D3S1358. The overall likelihood ratios for H_p versus H_{d1} are much larger than those for H_p versus H_{d2} . The captured screen for the competing hypotheses H_p versus H_{d2} is displayed in Figure 5.4.

5.8 Complex kinship determinations: method and software

All of the above discussions on paternity and kinship problems consider, at most, three persons. Analytical formulas are derived and computer programs have been developed. In practice, we also encounter complicated paternity problems which are difficult to handle analytically (e.g. a case in which the alleged father cannot be typed but several of his relatives can), complex and missing person problems in which relatives of the missing persons are typed.

Table 5.14 Likelihood ratios for paternity testing problems with hypotheses given in (5.19) and genotypes in Table 5.11, where Y is the child, X is the half-brother of the mother of the child and Z is the alleged father.

Locus	H_p versus H_{d1}		H_p versus H_{d2}	
	$\theta = 0$	0.01	0	0.01
D3S1358	1.26	1.27	1.11	1.11
vWA	6.25	5.35	1.72	1.68
FGA	3.66	3.06	1.57	1.51
Overall	28.82	20.79	3.00	2.84

Kinship of Three Persons

Load frequency file: _____

☒ theta = 0.01

Locus *FGA*

Y _____

X _____

Z _____

X and Z must be biologically unrelated

Calculate Save Clear Overall LR 3

Locus	Y	X	Z	LR	LR(0.01)
D3S1358	15,17	15,16	17,18	1.11	1.12
vWA	18,18	18,19	18,18	1.72	1.68
FGA	20,21	20,21	19,20	1.57	1.51

Figure 5.4 Captured screen with hypotheses H_p versus H_{d2} given in (5.19) and genotypes in Table 5.11, where Y is the child, X is the half-brother of the mother of the child and Z is the alleged father.

Dawid *et al.* (2002) and Egeland *et al.* (1997), respectively, used the probabilistic expert system and an algorithm resembling Elston–Stewart to handle complex pedigrees. There have also been discussions on the theory on general kinship determinations, with computer software developed accordingly (Brenner 1997; Egeland *et al.* 1997; Fung 2000, 2003b). The EasyDNA_In_1_Minute (Fung 2000, 2003b) is one such software that is easy to use.

The EasyDNA_In_1_Minute is a package consisting of four computer modules or programs developed for handling paternity and kinship determinations, including the statistical calculations for (a) alleged fathers, (b) alleged fathers where DNA typing is absent, (c) missing persons, and (d) incest cases. The programs can deal with both civil and criminal paternity cases. Computer enumeration is used in the calculations for complex paternity and kinship problems. The programs employ the pedigree tree design, and so are very easy to understand and use. The developed programs have wide applicability, for example the program EasyPA_In_1_Minute can handle the calculations both for the standard trio case and for the motherless paternity case, with or without DNA typing for the relatives of the mother. The program can deal with problems with more than one relative typed, and handle non-standard alternative hypotheses.

In the following sections, we describe the main features of the programs and explain the background theory and methods. We assume Hardy–Weinberg and linkage equilibrium, which

is commonly taken in paternity testing and kinship determinations (Brenner 1997; Dawid *et al.* 2002; Egeland *et al.* 1997, 2000; Fung *et al.* 2002, 1996; Gaytmenn *et al.* 2002; Lee *et al.* 1999; Thomson *et al.* 2001).

5.8.1 EasyPA_In_1_Minute software and the method

The EasyPA_In_1_Minute software can deal with the paternity testing problems that we have discussed in earlier sections and Chapter 4 when the population is taken to be in Hardy–Weinberg equilibrium. In addition, it can deal with problems in which many relatives are typed, e.g. mother not available but her relatives are. We are going to describe the method that the software is built upon, based on the motherless case in which relatives of the mother (M) provide the genetic information. Consider the usual hypotheses H_p : the alleged father (AF) is the true father (TF) of the child (C), and H_d : the true father is a random unrelated man. Suppose that the mother of mother (MoM) and the father of mother (FoM) have genotypes 16/17 and 17/19 at D3S1358, respectively (see Table 5.15). We are able to infer from the genotypes of MoM and FoM that the genotype of M is either 16/17, 16/19, 17/17 or 17/19. Given that the child's genotype is 17/18, the mother's genotype cannot be 16/19. The paternity index in this case can then be obtained by considering all of the mother's possible genotypes. As in D3S1358, we can infer that the genotype of the mother at vWA is either 17/19 or 17/20. So, the paternity index for vWA can be obtained similarly, noticing that the child's genotype is 17/19. Furthermore, the paternity index for FGA can also be derived in a similar way.

The EasyPA_In_1_Minute software can be run easily, since the pedigree tree diagram is shown. First, we need to import the allele frequency file (HKChinese.af) and the genotype file (GenotypePA.txt) to the program. Then, we select the names of the child, alleged father and relatives of the mother in the genotype file (in the file, they are called Name- C , Name- AF , Name- FoM and Name- MoM , and the corresponding names are C , AF , FoM and MoM in this particular example) using the built-in combo box. The paternity index for each locus and the overall paternity index are calculated and displayed immediately after the *Calculate* button is clicked. In other words, the paternity indices at *all loci* can be obtained immediately by clicking a few buttons. The paternity and kinship problem can be solved *within one minute*, no matter how many loci one has in the battery of tests. The captured screen in Figure 5.5 shows the details of the results. Another useful feature of the program is that the input names and the output findings can be saved in a file, which can be used for checking and/or reporting purposes. The findings for the above problem obtained by the program are summarized in the fourth column of Table 5.16.

In the above situation, genotypes of both parents of M are available, from which we can derive possible genotypes of M . This makes the calculations easy to handle. However, the situation becomes more complicated when the genotypes of neither parent, or only one parent,

Table 5.15 Genotype data of the child (C), alleged father (AF) and relatives (FoM , MoM , $S1oM$, $S2oM$) of the mother.

Locus	C	AF	FoM	MoM	$S1oM$	$S2oM$
D3S1358	17/18	18/20	16/17	17/19	17/17	17/17
vWA	17/19	19/19	17/17	19/20	17/19	17/19
FGA	20/22	22/23	20/21	22/23	20/22	21/23

are available. We consider below the situations in which only the siblings of M ($S1oM$) and/or the father of M are available for typing. Table 5.16 gives the paternity indices obtained by the EasyPA_In_1_Minute for various relatives combinations (see Table 5.15 for the genotypes). A method that the EasyPA_In_1_Minute or the general EasyDNA_In_1_Minute is based upon is to evaluate from the available genotype information of relatives the possible genotype(s) of the parent(s) of M by enumeration, which may possibly transmit on to M and then to C . The law of total probability, Bayes Theorem and conditional probability formulas are employed for assessing the probabilities of having those possible genotypes. For example, in the C - FoM - $S1oM$ - AF case at D3S1358, we can use the genotypes of FoM , $S1oM$ and C (see Table 5.15) to infer the genotype of MoM , which must be $17/y$, where y can be any allele at D3S1358. As a result, the genotype of M can be either one of $16/17$, $17/17$, $16/y$ or $17/y$. From this information, and using the conditional probability of observing each of these possible genotypes, we can compute the paternity index based on all possible values of y . Figure 5.6 illustrates some of the ideas which can be generalized to deal with other situations.

Suppose that we have another sibling of M ($S2oM$) typed, and he/she has the same genotype $17/17$ at D3S1358 as the first sibling ($S1oM$). In this case, $S2oM$ does not provide extra genotype information that the possible genotype of M is still $16/17$, $17/17$, $16/y$ or $17/y$. This may give rise to the assumption that the paternity index will remain unchanged. In fact this assumption is incorrect, because the conditional probability of observing a possible

Calculate Paternity Index (alleged father typed)

True Father

Mother

Child

NAME-C

Father of Mother

NAME-FOM

Mother of Mother

NAME-MOM

Sibling of Mother

H0: Alleged Father (AF)

NAME-AF

H1: True Father

A random man

Calculate

Reset

Paternity Index

D3S1358: 8.928571

VWA: 4.179320

FGA: 2.247757

Overall: 83.687897

Prior Probability of Paternity

0.5

Generate Output

Exit

Figure 5.5 Captured screen for running the EasyPA_In_1_Minute software for paternity testing in a motherless case with relatives of mother typed. H_0 and H_1 are used in the software to represent H_p and H_d .

Table 5.16 Paternity indices for H_p : TF of C is AF versus (a) H_d : TF of C is a random man, or (b) H_d : TF of C is a brother of AF .

Locus	<i>C-FoM-</i> <i>S1oM-</i> <i>C-FoM-</i> <i>S1oM-</i> <i>C-FoM-</i> <i>S2oM-</i> <i>C-S1oM-</i>						
	<i>C-M-AF</i>	<i>C-AF</i>	<i>MoM-AF</i>	<i>AF</i>	<i>S1oM-AF</i>	<i>S2oM-AF</i>	<i>C-S1oM-AF</i>
(a) H_d : TF of C is a random man							
D3S1358	8.93	4.46	8.93	6.75	8.07	8.26	7.48
vWA	9.42	4.71	4.18	7.79	4.24	4.04	3.16
FGA	2.25	1.40	2.25	2.59	2.19	2.25	2.14
Overall	18.9	29.5	83.9	136	74.9	75.0	50.6
(b) H_d : TF of C is a brother of AF							
D3S1358	1.80	1.63	1.80	1.74	1.78	1.78	1.76
vWA	1.81	1.65	1.61	1.77	1.62	1.61	1.52
FGA	1.38	1.17	1.38	1.44	1.37	1.38	1.36
Overall	4.50	3.15	4.02	4.46	3.95	3.96	3.65

genotype is no longer the same as before. The conditional probabilities in the two situations are

$$P(M|C, FoM, S1oM, AF) \text{ and } P(M|C, FoM, S1oM, S2oM, AF),$$

which are clearly not the same. Table 5.16 summarizes the paternity indices for various cases involving different relative combinations.

Another method that the program employs is by means of the law of total probability. To illustrate, we let the genotypes of the MoM and $S1oM$ (suppose that no other siblings of the

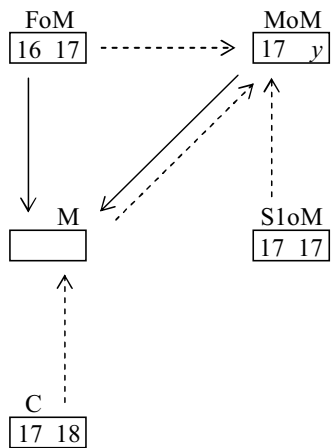


Figure 5.6 A diagram to illustrate some of the ideas used in the calculations in the EasyDNA_In_1_Minute software.

mother are available) be (x, y) and (u, v) , where x, y, u and v may take any possible allelic values. By the law of total probability in Equation (2.19), we obtain $P(x, y) = \sum_{u,v} P(x, y, u, v)$. The genotype probability of MoM can be obtained from the joint genotype probabilities $P(x, y, u, v)$'s, of which many are zeros, since MoM and $S1oM$ have at least one common allele.

5.8.2 EasyPant_In_1_Minute

When the alleged father (AF) is not available, his relatives may have to be typed. This is by no means a rare occurrence, and frequently arises in inheritance disputes, where the AF is deceased. Although there are methods provided in Chapter 4 and earlier sections of Chapter 5, they can only handle the situation with only one relative of AF typed. A computer program EasyPant_In_1_Minute with a more complicated pedigree diagram has been developed for this purpose. Previous ideas in Section 5.8.1 can be generalized, though the theory here would be more involved. For example, consider the case in which the AF is not available and siblings 1 and 2 of the AF are typed with genotypes 19/20 and 16/19 at D3S1358, respectively. Suppose that the M is not available but the FoM and MoM are. In this case, neither the M nor the AF can be typed, and, instead, two relatives for each of them are typed. A total of five persons (including the C) are involved. Relatively little attention has been paid so far in the literature to complex paternity cases of this kind, except Brenner (1997), Egeland *et al.* (2000) and Fung (2003b). Based on the software, we find that the paternity index of this $C-FoM-MoM-S1oAF-S2oAF$ case is 0.404, which is much smaller than the paternity index, 8.93, of the $C-FoM-MoM-AF$ case that the AF is typed (Table 5.16). At this particular locus D3S1358, the new paternity index is smaller than 1 and so it does not seem to support the null hypothesis that the AF (who is not typed) is the TF of the C . Such a low value is, however, not surprising, since the two siblings of the AF do not even have allele 18, which the C has inherited from his TF .

5.8.3 EasyIN_In_1_Minute

Some researchers find it hard to deal with incest cases because the AF and the M are biologically related, but, in fact, these cases are usually not difficult to handle (see Section 4.2.1). Consider a criminal paternity incest case with the genotype information given in Table 5.17. A child ('child 1 of F and M ', called Name-C1 in the genotype file) and his mother are accused of having an incestuous relationship, from which the mother has given birth to another child (C). The hypotheses of interest are

H_p : the TF of C is 'child 1 of F and M ';

H_d : the TF of C is a random unrelated man.

(Notice that H_0 and H_1 , instead of H_p and H_d , are used in the software.) This is not a difficult problem and our program finds the overall paternity index at all three loci equal to 13.4 (details omitted). Suppose that the accused puts up an alternative explanation H_{d1} : the TF of C is F , and there is no incestuous relationship. How should we compute the paternity index?

If the genotype of F is available, the problem can be easily solved. If it is not, the problem is non-trivial and not many discussions are observed. Consider the case in which the genotype of F is unavailable, but the genotype of his two siblings and the M are. (They are called Name-S1oF, Name-S2oF and Name-M in the genotype file, for genotypes; see Table 5.17.)

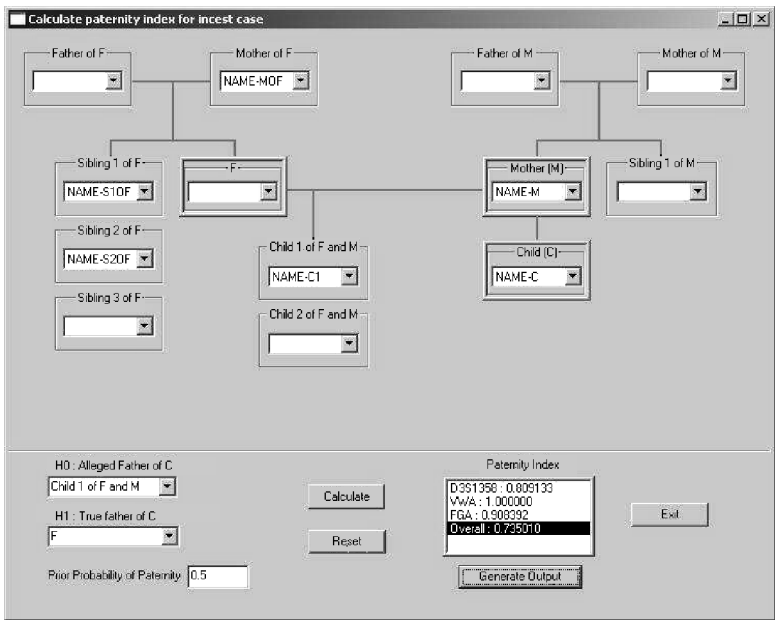


Figure 5.7 Captured screen for running the EasyIN_In_1_Minute software for an incest case with a special alternative explanation $H1$ (i.e. H_d).

Based on the ideas of pedigree analysis, the law of total probability and the Bayes formula, the EasyIN_In_1_Minute can solve this problem by just clicking in the available genotypes information. By choosing the appropriate option in the combo box $H1$ in the software (meaning H_d here), the paternity index can be obtained immediately. The paternity index in this case is reduced from 13.4 to 0.735 (see Figure 5.7). The index, which is even smaller than 1 may suggest that the genetic evidence is more in favor of the accused. Of course, the genetic evidence at other loci as well as the non-genetic evidence have to be considered in order to come up with a more definite answer.

5.8.4 EasyMISS_In_1_Minute

Suppose that a person (X) went missing and his family members reported his disappearance to local police. A dead body (Alleged X , abbreviated as AX) was found a few months later. In order to determine if the dead body is X or not, we type a total of eight family members

Table 5.17 An incest case in which ‘Child 1 of F and M ’ and his mother (M) are accused of having an incestuous relationship with genotype data of C , M , Child 1 and relatives (MoF , $S1oF$ and $S2oF$) of F .

Locus	C	M	Child 1	MoF	$S1oF$	$S2oF$
D3S1358	16/18	17/18	16/18	15/16	15/18	16/18
vWA	17/18	17/17	17/18	15/19	15/19	18/19
FGA	21/22	22/23	21/22	21/23	23/24	21/24

Calculate likelihood ratio for missing person

Father of X: [NAME-FIL]
 Mother of X: [NAME-MIL]
 Sibling of X: [NAME-S1]
 Sibling of X: [NAME-S2]
 Sibling of X: [NAME-S3]
 Sibling of X: [NAME-S4]
 Missing Person: [NAME-MP]
 S (Spouse of X): [NAME-SIL1]
 Sibling of S: [NAME-SIL2]
 Sibling of S: [NAME-SIL3]
 Sibling of S: [NAME-SIL4]
 Child: [NAME-C1]
 Child: [NAME-C2]
 Child: [NAME-C3]
 Alleged X: [NAME-AX]
 Prior Probability of Paternity: 0.5

Calculate
 Reset
 Exit
 Generate Output

Likelihood Ratio

CSF1PO	: 5.292013
D5S818	: 68.027211
D13S317	: 5.047064
D7S820	: 664.893617
Overall	: 8.873351e+013

Figure 5.8 Captured screen for running the EasyMISS_In_1_Minute software for a missing person problem with genotypes of eight family members provided in Table 5.18.

of X , i.e., his three sibs (genotype data labeled as $S1$, $S2$, $S3$), his two children ($C1$, $C2$), his father-in-law (FIL) and two sisters-in-law ($SIL1$, $SIL2$). These genotype data at nine loci obtained from the AmpF ℓ STR Profiler kit are put in a file (Missing.txt) as shown in Table 5.18. We are interested in the following hypotheses:

H_p : the dead body (Alleged X) is the missing person X ;

H_d : the dead body is an unrelated random man.

Although the problem looks complicated, the computer program EasyMISS_In_1_Minute can deal with it easily. We first import the allele frequency file (HKChinese.af) and the genotype file (Missing.txt). Then we select the names of Alleged X and the eight family members in the genotype file using the built-in combo boxes. The likelihood ratio at each locus and the overall likelihood ratio are calculated and displayed almost instantly when we click the *Calculate* button (see the captured screen shown in Figure 5.8 for details).

Figure 5.9 shows the captured screen for the output file, in which we can check whether we have input correctly. Notice that in the EasyDNA_In_1_Minute software, there would be little manual handling error, since only very minimal information is needed for the input. Moreover, the findings can be obtained quickly, *within one minute*.

Tested Person						Sample Info	
AX	Alleged X					NAME-AX	
XS1	Sibling 1 of Father					NAME-S1	
XS2	Sibling 2 of Father					NAME-S2	
XS3	Sibling 3 of Father					NAME-S3	
SP	Father of Spouse					NAME-FIL	
SS1	Sibling 1 of Spouse					NAME-SIL1	
SS2	Sibling 2 of Spouse					NAME-SIL2	
C1	Child 1 of X					NAME-C1	
C2	Child 2 of X					NAME-C2	

Overall Summary					
Locus tested	AX	XS1	XS2	XS3	SP
D3S1358	15/17	15/17	18/17	18/19	14/16
VWA	20/16	14/16	20/19	14/19	17/18
FGA	24/19	21/19	21/25.2	24/19	23/23
TH01	6/9.3	6/9.3	11/9	6/9	9/9
TPOX	11/10	8/10	11/10	8/9	8/8
CSF1P0	11/12	11/10	12/12	12/10	9/11
D5S818	12/7	13/13	12/7	12/13	12/11
D13S317	8/8	8/12	11/12	11/8	11/9
D7S820	11/7	11/7	8/10	11/10	12/12

Locus tested	SS1	SS2	C1	C2	LR
D3S1358	16/17	14/17	15/14	15/16	3.1602
VWA	17/16	18/18	20/17	16/16	160.2564
FGA	23/25	23/26.2	19/26.2	24/23	46.3392
TH01	9/9	9/7	9.3/9	9.3/7	86.2069
TPOX	8/11	8/12	10/8	10/8	36.3055
CSF1P0	9/10	9/15	11/11	12/11	5.2920
D5S818	11/10	12/10	7/11	12/10	68.0272
D13S317	11/9	9/10	8/11	8/9	5.0471
D7S820	12/11	12/11	7/12	11/13	664.8936
Overall					8.87e+013

Figure 5.9 Captured screen for the output file of the missing person example.

5.8.5 Other considerations: probability of paternity and mutation

Another feature found in the EasyDNA_In_1_Minute software is that prior probabilities that the prosecution proposition H_p holds true have been built in. Besides, there is an option to choose your own prior; see, for example, the lower left boxes in Figures 5.7 and 5.8. The posterior probabilities or probabilities of paternity are automatically evaluated after obtaining the likelihood ratio/paternity index for the problem of kinship determination.

In the case of paternity testing, there is a possibility that the genotypes of all except one or two loci of the AF and C are found to match one another. This phenomenon may be explained

Table 5.18 Names and genotype data of nine persons for a missing person example, from Fung *et al.* (2006). (Reproduced by permission of Elsevier.)

Sample	D3S1358	vWA	FGA	THO1	TPOX	CSF1PO	D5S818	D13S317	D7S820
AX	15/17	20/16	24/19	6/9.3	11/10	11/12	12/7	8/8	11/7
S1	15/17	14/16	21/19	6/9.3	8/10	11/10	13/13	8/12	11/7
S2	18/17	20/19	21/25.2	11/9	11/10	12/12	12/7	11/12	8/10
S3	18/19	14/19	24/19	6/9	8/9	12/10	12/13	11/8	11/10
C1	15/14	20/17	19/26.2	9.3/9	10/8	11/11	7/11	8/11	7/12
C2	15/16	16/16	24/23	9.3/7	10/8	12/11	12/10	8/9	11/13
FIL	14/16	17/18	23/23	9/9	8/8	9/11	12/11	11/9	12/12
SIL1	16/17	17/16	23/25	9/9	8/11	9/10	11/10	11/9	12/11
SIL2	14/17	18/18	23/26.2	9/7	8/12	9/15	12/10	9/10	12/11

by mutation. The EasyPA_In_1_Minute program also allows the possibility of mutation and the simple average mutation paternity index (Section 4.10.2) is adopted:

$$AMPI = \frac{\text{average mutation rate (AMR)}}{\text{power of exclusion (PE)}}$$

for the mismatch locus. A standard trio case example with one mismatch locus is given in Section 4.10.

5.9 Problems

- Let $X = A_1A_2$ and $Y = A_2A_2$, with $p_1 = 0.12$ and $p_2 = 0.37$. Test whether X and Y are first cousins versus whether they are unrelated, and obtain its likelihood ratio. The population is in Hardy–Weinberg equilibrium.
- In order to test whether Y is the nephew or unrelated to X , we type Y and a son of X . The DNA profiles of Y and Z (the son of X) are A_1A_2 and A_1A_3 , respectively. Obtain the likelihood ratio for the kinship determination. Hardy–Weinberg law is assumed.
- In a subdivided population with the degree of subdivision θ , test whether $X = A_1A_2$ and $Y = A_1A_2$ are parent–child versus whether they are full siblings, and obtain its likelihood ratio. (Note: this situation may happen in immigration applications.) Evaluate the likelihood ratio when $p_1 = 0.18$, $p_2 = 0.43$ and $\theta = 0.03$.
- In a subdivided population with the degree of subdivision θ , find the paternity index for a standard trio case in which the genotypes of the child, mother and alleged father are $C = A_1A_2$, $M = A_1A_1$ and $AF = A_2A_3$, respectively, and the two competing hypotheses are

H_p : the alleged father is the true father of the child;
 H_d : the alleged father and the child are biologically unrelated.

 Evaluate the paternity index when $p_1 = 0.18$, $p_2 = 0.23$, $p_3 = 0.31$ and $\theta = 0.03$.
- In a paternity test case, the genotypes of the child, mother and alleged father are A_1A_1 , A_1A_2 and A_1A_3 , respectively. Find the paternity index about the following two competing

hypotheses:

H_p : the alleged father is the true father of the child;

H_d : a full sibling of the alleged father is the true father of the child.

The population is subdivided with the degree of subdivision θ . Evaluate the paternity index when $p_1 = 0.12$, $p_2 = 0.23$, $p_3 = 0.38$ and $\theta = 0.03$.

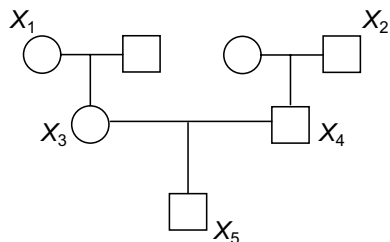
6. In a paternity testing case, the alleged father is not available and a full sibling R of the alleged father is typed instead. Suppose the genotypes of the child, mother and R are A_1A_1 , A_1A_1 and A_1A_2 , respectively. All involved people come from the same subdivided population with the degree of subdivision θ . Find the likelihood ratio about the following two competing hypotheses:

H_p : the alleged father, a brother of R , is the true father of the child;

H_d : a random unrelated man is the true father of the child.

7. Suppose $X = A_1A_1$, $Y = A_1A_2$ and $Z = A_2A_3$. Test whether X is the maternal uncle of Y and Z is the paternal grandfather of Y versus X , Y and Z are biologically unrelated. The population is in Hardy–Weinberg equilibrium.

8. Let $X_1 = A_1A_2$, $X_2 = A_3A_3$, $X_3 = A_1A_4$, $X_4 = A_2A_3$ and $X_5 = A_1A_3$. The family relationship among X_1 , X_2 , X_3 , X_4 and X_5 is shown in the figure on the right. Find $P(X_1, X_2, X_3, X_4, X_5)$ and $P(X_1, X_2, X_3, X_5)$ if X_4 is unavailable. The population to which the family belongs is in Hardy–Weinberg equilibrium.



9. In order to determine the relationship between two individuals Y and Z , their genotypes at three loci are provided as follows: at locus D3S1358, $Y = 13/13$, $Z = 13/14$; at locus vWA, $Y = 16/18$, $Z = 16/18$; at locus FGA, $Y = 19/19$, $Z = 20/21$. The following three hypotheses are of interest:

H_{p1} : Y and Z are full siblings;

H_{p2} : Y and Z are half siblings;

H_d : Y and Z are biologically unrelated.

The allele frequencies are listed in Table 4.3. For $\theta = 0$ and 0.03 , test H_{p1} versus H_d , and H_{p2} versus H_d .

- Use the computer program EasyDNA_2Persons to obtain the likelihood ratios at these three loci.
- Use Table 5.5 to evaluate the likelihood ratios at these three loci.
- Check whether the corresponding results in (a) and (b) are the same.