Trio detection
Jonathan Bloom
jbloom@broadinstitute.org
July 10, 2014

Let $\theta$ and $\varphi$ denote the haplotypes of samples $s_\theta$ and $s_\varphi$ producing sequencing data sets $x$ and $y$, respectively. Let $r$ be a categorical random variable taking values in a set $R$ of potential relationships between the samples. For this note, we’ll set

$$R = \{ \text{identical, parent, grandparent, sibling, aunt/uncle, unrelated} \}$$

where an element completes the sentence

“Sample $s_\varphi$ is __ to/of Sample $s_\theta$."

Then under the simplifying assumption that $R$ accounts for all possible relationships, we have

$$p(r \mid x, y) = \frac{p(x, y \mid r) p(r)}{\sum_r p(x, y \mid r) p(r)},$$

(1)

Even without this assumption, we can compute the posterior odds of relationship $r_1$ versus $r_2$ as

$$\frac{p(r_1 \mid x, y)}{p(r_2 \mid x, y)} = \frac{p(x, y \mid r_1)}{p(x, y \mid r_2)} \cdot \frac{p(r_1)}{p(r_2)}.$$  (2)

The likelihood $p(x, y \mid r)$ expands as

$$p(x, y \mid r) = \sum_{\theta, \varphi} p(x, y \mid \theta, \varphi, r) p(\theta, \varphi \mid r)$$

$$= \sum_{\theta, \varphi} p(x \mid \theta) p(y \mid \varphi) p(\theta \mid \varphi, r) p(\varphi).$$

(3)

As show in Proposition 1 below, the term $p(\theta \mid \varphi, r)$ is determined by the haplotype distribution $H$ and Mendelian inheritance array $M$:

$$H_\psi \equiv p(\psi),$$

$$M_\theta^{\varphi, \psi} \equiv p(\theta \mid (s_\varphi, s_\psi) \rightarrow s_\theta, \varphi, \psi).$$

(5)

(6)
Here we use → as shorthand for “parent of,” so \((s_\varphi, s_\psi) \to s_\theta\) denotes the event that samples \(s_\varphi\) and \(s_\psi\) are parents of sample \(s_\theta\). \(H\) and \(M\) also determine the array \(N\) encoding the joint probability of parent haplotypes \((\varphi, \psi)\) given offspring haplotype \(\theta\):

\[
N^\theta_{\varphi,\psi} \equiv p(\varphi, \psi \mid (s_\varphi, s_\psi) \to s_\theta, \theta) = \frac{p(\theta, \varphi, \psi \mid (s_\varphi, s_\psi) \to s_\theta)}{p(\varphi, \psi \mid (s_\varphi, s_\psi) \to s_\theta)} \cdot \frac{p(\varphi, \psi, \theta \mid (s_\varphi, s_\psi) \to s_\theta)}{p(\theta \mid (s_\varphi, s_\psi) \to s_\theta)} = M^\varphi,\psi_\theta \cdot \frac{H_\varphi H_\psi}{H_\theta}.
\]

Finally, ordering the haplotypes (AA, AB, BB), the matrices \(M_\theta\) are:

\[
M_{\text{AA}} = \begin{pmatrix} 1 & .5 & 0 \\ .5 & .25 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad M_{\text{AB}} = \begin{pmatrix} 0 & .5 & 1 \\ .5 & .5 & .5 \\ 1 & .5 & 0 \end{pmatrix} \quad M_{\text{BB}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & .25 & .5 \\ 0 & .5 & 1 \end{pmatrix}
\]

**Proposition 1.** Considering each relationship in turn, we have:

\[
p(\theta \mid \varphi, r = \text{identical}) = \begin{cases} H_\varphi, & \text{if } \theta = \varphi \\ 0, & \text{otherwise} \end{cases} \equiv P^\varphi_
\]

\[
p(\theta \mid \varphi, r = \text{parent}) = \sum_{\psi} M^\varphi,\psi_\theta H_\psi \equiv P^\varphi_
\]

\[
p(\theta \mid \varphi, r = \text{grandparent}) = \sum_{\psi} P^\psi_\theta P^\varphi_\psi \equiv S^\varphi_
\]

\[
p(\theta \mid \varphi, r = \text{sibling}) = \sum_{\psi,\psi'} M^\psi,\psi'_\theta N^\varphi_{\psi,\psi'} \equiv S^\varphi_
\]

\[
p(\theta \mid \varphi, r = \text{aunt/uncle}) = \sum_{\psi} P^\psi_\theta S^\varphi_\psi \equiv S^\varphi_
\]

\[
p(\theta \mid \varphi, r = \text{unrelated}) = H_\theta
\]

**Proof.** The event \((r = \text{identical})\) is equivalent to \(s_\theta = s_\varphi\).

\[
p(\theta \mid \varphi, s_\theta = s_\varphi) = \begin{cases} p(\varphi), & \text{if } \theta = \varphi \\ 0, & \text{otherwise} \end{cases} = \begin{cases} H_\varphi, & \text{if } \theta = \varphi \\ 0, & \text{otherwise} \end{cases}
\]
The event \((r = \text{parent})\) is equivalent to \(s_\varphi \rightarrow s_\theta\).

\[
p(\theta \mid \varphi, s_\varphi \rightarrow s_\theta) = \sum_\psi p(\theta \mid \varphi, \psi, (s_\varphi, s_\psi) \rightarrow s_\theta)p(\psi \mid \varphi, (s_\varphi, s_\psi) \rightarrow s_\theta)
= \sum_\psi p(\theta \mid \varphi, (s_\varphi, s_\psi) \rightarrow s_\theta)p(\psi)
= \sum_\psi M_{\theta,\psi}^\varphi H_\psi
\]

The event \((r = \text{grandparent})\) is equivalent to \((s_\varphi \rightarrow s_\psi, s_\psi \rightarrow s_\theta)\) where \(\psi\) is the haplotype of the intermediate parent \(s_\psi\).

\[
p(\theta \mid \varphi, s_\varphi \rightarrow s_\psi, s_\psi \rightarrow s_\theta) = \sum_\psi p(\theta \mid \psi, \varphi, s_\varphi \rightarrow s_\psi, s_\psi \rightarrow s_\theta)p(\psi \mid \varphi, s_\varphi \rightarrow s_\psi, s_\psi \rightarrow s_\theta)
= \sum_\psi p(\theta \mid \psi, s_\varphi \rightarrow s_\theta)p(\psi \mid \varphi, s_\varphi \rightarrow s_\psi)
= \sum_\psi P_{\theta,\psi}^\varphi P_\psi^{s_\varphi}
\]

The event \((r = \text{siblings})\) is equivalent to \((s_\varphi, s_\psi, s_\psi' \rightarrow (s_\varphi, s_\theta))\) where \(\psi\) and \(\psi'\) are the haplotypes of the common parents \(s_\psi\) and \(s_\psi'\). We also use \(\leftrightarrow\) as shorthand for “sibling of.”

\[
p(\theta \mid \varphi, s_\varphi \leftrightarrow s_\theta) = \sum_{\psi,\psi'} p(\theta \mid \psi, \psi', \varphi, (s_\psi, s_\psi') \rightarrow (s_\varphi, s_\theta))p(\psi, \psi' \mid \varphi, (s_\psi, s_\psi') \rightarrow (s_\varphi, s_\theta))
= \sum_{\psi,\psi'} p(\theta \mid \psi, \psi', (s_\psi, s_\psi') \rightarrow s_\theta)p(\psi, \psi' \mid \varphi, (s_\psi, s_\psi') \rightarrow s_\varphi)
= \sum_{\psi,\psi'} M_{\theta,\psi',\psi}^{s_\varphi} N_{\psi,\psi'}^{s_\varphi}
\]

The event \((r = \text{aunt/uncle})\) is equivalent to \((s_\varphi \leftrightarrow s_\psi, s_\psi \rightarrow s_\theta)\) where \(\psi\) is the haplotype of the sibling \(s_\psi\) of \(s_\varphi\) who is also a parent of \(s_\theta\).

\[
p(\theta \mid \varphi, s_\varphi \leftrightarrow s_\psi, s_\psi \rightarrow s_\theta) = \sum_\psi p(\theta \mid \psi, \varphi, s_\varphi \leftrightarrow s_\psi, s_\psi \rightarrow s_\theta)p(\psi \mid \varphi, s_\varphi \leftrightarrow s_\psi, s_\psi \rightarrow s_\theta)
= \sum_\psi p(\theta \mid \psi, s_\varphi \rightarrow s_\theta)p(\psi \mid \varphi, s_\varphi \leftrightarrow s_\psi)
= \sum_\psi P_{\theta,\psi}^s S_{\psi}^{s_\varphi}
\]
The haplotypes $\theta$ and $\varphi$ are conditionally independent given ($r = \text{unrelated}$).

$$p(\theta \mid \varphi, r = \text{unrelated}) = p(\theta) = H_{\theta}$$

Suppose there is a de novo mutation probability of $\delta$ between major and minor alleles. Then, ordering the haplotypes (AA, AB, BB), the probability $D_{\theta'}^{\theta}$ of mutation from $\theta'$ to $\theta$ is given in row $\theta$ and column $\theta'$ of the matrix

$$D \equiv \begin{pmatrix}
(1 - \delta)^2 & \delta(1 - \delta) & \delta^2 \\
2\delta(1 - \delta) & \delta^2 + (1 - \delta)^2 & 2\delta(1 - \delta) \\
\delta^2 & \delta(1 - \delta) & (1 - \delta)^2
\end{pmatrix} \quad (7)$$

To incorporate mutation in Proposition 1, replace the Mendelian inheritance array $M$ in (6) by the inherence-plus-mutation array $\tilde{M}$ defined by

$$\tilde{M}_{\theta}^{\varphi, \psi} \equiv \sum_{\theta'} D_{\theta}^{\theta'} M_{\theta'}^{\varphi, \psi} \quad (8)$$

and similarly replace $N_{\varphi, \psi}^{\theta}$ by

$$\tilde{N}_{\varphi, \psi}^{\theta} \equiv \tilde{M}_{\theta}^{\varphi, \psi} \cdot \frac{H_{\varphi} H_{\psi}}{H_{\theta}}. \quad (9)$$