Probability of sharing a rare variant by related individuals

Alexandre Bureau[1,2], Ingo Ruczinski[3]

Received _____; accepted _____

[1]Centre de recherche de l'Institut universitaire en santé mentale de Québec

[2]Département de médecine sociale et préventive, Université Laval

[3]Johns Hopkins Bloomberg School of Public Health

1. Methods

Our goal is to compute the probability that a set of related subjects whose DNA sequence is observed through sequencing or other means (sequenced subjects) share a rare variant (RV) identical-by-descent given that a RV has been observed at a site in the sequence. We assume that the variant for which we compute a sharing probability is rare enough that there exists a single copy of that variant among the alleles present in the n_f founders of the pedigree relating the subject for which we want to compute a sharing probability. In the basic setting, all founders are unrelated and a single copy of the variant is present among the founders. In a generalization, we later allow founders to be related, and two copies of the allele to be introduced in the pedigree by a pair of related founders.

1.1. Computation assuming all founders are unrelated

We define the following random variables and constants:

C_i Number of copies of the RV received by sequenced subject i

F_j Indicator variable that founder j introduces one copy of the RV in the pedigree

B_k Number of copies of the RV in subject k where a line of descent from a founder branches into two separate lines of descents to a subset of sequenced subjects

D_{ij} Number of generations (meioses) between subject i and his ancestor j

For a set of n sequenced subjects for which the pedigree structure limits to one the number of copies of the rare variant that they can share, we want to compute the probability

$$\begin{aligned} P[\text{RV shared}] &= P[C_1 = \ldots = C_n = 1 | C_1 + \ldots + C_n \geq 1] = \frac{P[C_1 = \ldots = C_n = 1]}{P[C_1 + \ldots + C_n \geq 1]} \quad (1) \\ &= \frac{\sum_{j=1}^{n_f} P[C_1 = \ldots = C_n = 1 | F_j] P[F_j]}{\sum_{j=1}^{n_f} P[C_1 + \ldots + C_n \geq 1 | F_j] P[F_j]} \end{aligned}$$

where the expression on the second line results from our assumption that there exists a single copy of that variant among the alleles present in the n_f founders. The probabilities $P[F_j] = \frac{1}{n_f}$ cancel from the numerator and denominator. For the other terms, we first derive expressions for the special case where all the sequenced subjects descend from every founder among their ancestors through independent lines of descent. In that case,

$$P[C_1 = \ldots = C_n = 1 | F_j] = \begin{cases} \prod_i \left(\frac{1}{2}\right)^{D_{ij}} = \left(\frac{1}{2}\right)^{D_j} & \text{if } F_j \text{ is a common ancestor to the } n \text{ sequenced subjects} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

and

$$P[C_1 + \ldots + C_n \geq 1 | F_j] = 1 - P[C_1 = \ldots = C_n = 0 | F_j] = 1 - \prod_{i \in d(j)} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}}\right) \quad (3)$$

where $D_j = \sum_i D_{ij}$ and $d(j)$ is the subset of sequenced individuals who descend from founder j.

The global expression is then

$$P[\text{RV shared}] = \frac{\sum_{j=1}^{n_f} \left(\frac{1}{2}\right)^{D_j} I(F_j \text{ is a common ancestor to the } n \text{ sequenced subjects})}{\sum_{j=1}^{n_f} \left[1 - \prod_{i \in d(j)} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}}\right)\right]} \quad (4)$$

We note here that equation 4 covers the general case of pedigrees without inbreeding, including individuals marrying multiple times and marriage loops. For the common special case of a pedigree with a founder couple ancestral to all descendants in the pedigree, the

numerator simplifies and we obtain the following expression:

$$P[\text{RV shared}] = \frac{\left(\frac{1}{2}\right)^{D_f - 1}}{\sum_{j=1}^{n_f}\left[1 - \prod_{i \in d(j)}\left(1 - \left(\frac{1}{2}\right)^{D_{ij}}\right)\right]} \quad (5)$$

where f is any of the two founders forming the ancestral couple.

When the lineages of sequenced individuals "coalesce" at a branching individual k below a founder of the pedigree, we can no longer write a general expression like 4, and recursive computations are required. Without loss of generality, let k be the branching individual who has sequenced subjects $1, \ldots, i_k$ as descendants through independent lines of descent. We have

$$P[C_1 = \ldots = C_n = 1] = P[C_1 = \ldots = C_{i_k} = 1 | B_k = 1] P[B_k = C_{i_k+1} = \ldots = C_n = 1] \quad (6)$$

because $P[C_1 = \ldots = C_{i_k} = 1 | B_k = 0] = 0$. The term $P[C_1 = \ldots = C_{i_k} = 1 | B_k = 1]$ is computed from equation 2 replacing F_j by B_k. The term $P[B_k = C_{i_k+1} = \ldots = C_n = 1]$ is computed by reapplying equation 6 recursively with every branching individual.

Also, for a founder above a branching individual in the pedigree, we have

$$\begin{aligned} P[C_1 = \ldots = C_n = 0 | F_j] &= P[C_1 = \ldots = C_{i_k} = 0 | B_k = 1, F_j] P[B_k = 1, C_{i_k+1} = \ldots = C_n = 0 | F_j] \\ &\quad + P[C_1 = \ldots = C_{i_k} = 0 | B_k = 0, F_j] P[B_k = C_{i_k+1} = \ldots = C_n = 0 | F_j] \\ &= P[C_1 = \ldots = C_{i_k} = 0 | B_k = 1] P[B_k = 1, C_{i_k+1} = \ldots = C_n = 0 | F_j] \quad (7) \\ &\quad + P[B_k = C_{i_k+1} = \ldots = C_n = 0 | F_j] \end{aligned}$$

The term $P[C_1 = \ldots = C_{i_k} = 0 | B_k = 1]$ is computed from the right-hand side of equation 3 replacing F_j by B_k. The two terms $P[B_k = a, C_{i_k+1} = \ldots = C_n = 0 | F_j], a = 0, 1$ require recursive computations. If h is a branching individual who is an ancestor of k and a

descendant of founder j, then

$$\begin{aligned} P[B_k = 1, C_{i_k+1} = \ldots = C_n = 0|F_j] &= P[B_k = 1, C_{i_k+1} = \ldots = C_{i_h} = 0|B_h = 1]P[B_h = 1, C_{i_h+1} = \ldots = \\ &\quad + P[B_k = 1, C_{i_k+1} = \ldots = C_{i_h} = 0|B_h = 0]P[B_h = C_{i_h+1} = \ldots = \\ &= \left(\frac{1}{2}\right)^{D_{kh}} P[C_{i_k+1} = \ldots = C_{i_h} = 0|B_h = 1]P[B_h = 1, C_{i_h+1} = \ldots = \end{aligned}$$

and similarly to 8

$$\begin{aligned} P[B_k = C_{i_k+1} = \ldots = C_n = 0|F_j] &= \left(1 - \left(\frac{1}{2}\right)^{D_{kh}}\right) P[C_{i_k+1} = \ldots = C_{i_h} = 0|B_h = 1]P[B_h = 1, C_{i_h+1} = \\ &\quad + P[B_h = C_{i_h+1} = \ldots = C_n = 0|F_j] \end{aligned}$$

where the computation of the term $P[C_{i_k+1} = \ldots = C_{i_h} = 0|B_h = 1]$ can itself involve other branching individuals below h.

1.2. Computation allowing for relatedness between founders or inbreeding loops within a pedigree

We generalize our computation to the setting where founders are related, while still excluding that the founders are themselves inbred (only their children will be). This includes the setting where inbreeding loops are known and are included in the pedigree structure. One can then define a noninbred subpedigree by removing some familial links. The relatedness between the "founders" of that subpedigree can be captured by their kinship coefficient based on the removed links, and the first approximation described below can then be applied. When familial links between founders are unknown, they sometime can be estimated from genotype data on these founders. Other times, genotype data is only available on the sequenced subjects.

We propose two methods to approximate sharing probabilities between sequenced subjects in presence of IBD sharing in excess of what is expected based on the pedigree structure. With the first method, only one founder allele (not necessarily the RV considered in the computation) can be shared by only one pair of founders. This method gives an exact sharing probability when only two founders are related, and a good approximation when a few founders are related. Known founder pair-specific kinship coefficients can be used. With the second method, up to two alleles can be shared by two pairs of founders, and that number can be increased. It requires to assume that all founders are related to the same extent, i.e. all pairs of founders have the same kinship coefficient calculated to explain the excess sharing between sequenced subjects. The method gives a good approximation for more extensive hidden relatedness than the first method. Note that in this second approximation, we still assume that only two founders introduce a copy of the RV considered in the computation.

The elements that we need to implement either approach are:

1. The probability that a pair of related founders introduce the RV in the pedigree.
2. The sharing probabilities conditional on the introduction of the RV by two of the founders.

1.2.1. Probability that a pair of related founders introduce the RV in the pedigree

Method 1

The probability that two related founders, say j and k, introduce the RV in the pedigree is expressed as follows:

$$P[F_j, F_k] \;=\; P[\text{Allele shared is RV}|j\&k \text{ share allele IBD}]P[j\&k \text{ share allele IBD}] \quad (10)$$

$$= \frac{1}{2n_f - 1} 2\phi_{jk} = \frac{2\phi_{jk}}{2n_f - 1}$$

where ϕ_{jk} is the kinship coefficient between founders j and k. The first term represents the probability that the RV is the allele IBD between the two founders among the $2n_f - 1$ distinct alleles in all founders. The marginal probability that any founder h introduces the RV needs to be adjusted compared to the unrelated case. In that computation, we make the simplifying assumption that the probability that 3 or more founders share an allele IBD is 0 so that the event "i and j share an allele IBD" means that they are the only ones to do so. This assumption is true only when a single pair of founders are related. While the formula allows all pairs of founders to be related, we recommend using this approximation when only a few of the ϕ_{jk} are non-zero.

$$\begin{aligned} P[F_h] &= \sum_j \sum_{k>j} P[F_h | j\&k \text{ share allele IBD}] P[j\&k \text{ share allele IBD}] \\ &\quad + P[F_h | \text{no founder pair shares allele IBD}] P[\text{no founder pair shares allele IBD}] \\ &= \frac{2}{2n_f - 1} \sum_j \sum_{k>j} P[j\&k \text{ share allele IBD}] + \frac{1}{n_f} \left(1 - \sum_j \sum_{k>j} P[j\&k \text{ share allele IBD}] \right) \\ &= \frac{4 \sum_j \sum_{k>j} \phi_{jk}}{2n_f - 1} + \frac{1}{n_f} \left(1 - \sum_j \sum_{k>j} 2\phi_{jk} \right) \end{aligned} \tag{11}$$

We obtain the probability of F_j^U, the event that founder j is the only one to introduce the RV in the family, as

$$P[F_j^U] = P[F_j] - \sum_{k \neq j} P[F_j, F_k] \tag{12}$$

If we know which founders j and k are related, then their degree of relatedness is usually also known, and specifies their kinship coefficient ϕ_{jk}. If it is possible to identify a

subset of founders that are suspected to be related, with the other founders unrelated to that subset and between themselves, then this method can still be applied, with the kinship coefficient between the subset of founders suspected to be related estimated as described under Method 2. If instead familial links between founders are completely unknown, we generally recommend to apply the second method.

Method 2

For the second method, we assume $\phi_{jk} = \phi^f \forall j, k$. This is an assumption that we prefer to make when the relatedness between specific pairs of founders is unknown and we need to rely on genotye data to estimate it. Even with perfect information on IBD sharing between subjects, there is considerable variation in the kinship coefficient based on IBD sharing estimated for pairs of subjects with the same degree of relatedness due to variation in the length of genome shared from pair to pair (**?**), and reliable inference can only be obtained for the mean or other central tendency parameter.

Two situations can occur with respect to the genotype data available to estimate kinship between founders:

1. Polymorphic markers have been genotyped on the pedigree founders, typically a genomewide SNP array. Then ϕ_{jk} can be estimated for each founder pair j and k, and a global estimate $\hat{\phi}^f$ obtained by averaging the $\hat{\phi}_{jk}$ over all founder pairs from the same population.

2. Genotype data is only available on the sequenced subjects (either from the sequencing data itself of from other genotyping). The common ϕ^f is estimated based on the estimated kinship coefficients between sequenced subjects and the relationship

between the sequenced subjects and all founders.

$$\begin{aligned}\phi_{i_1 i_2} &= \phi^f \sum_j \sum_{k>j} \left[\left(\frac{1}{2}\right)^{D_{i_1 j}+D_{i_2 k}} I(j\&k \text{ not mating}) + \left(\frac{1}{2}\right)^{D_{i_1 j}+D_{i_2 k}-1} I(j\&k \text{ mating}) \right] + \phi^p_{i_1 i_2} \\ &= \phi^f \kappa_{i_1 i_2} + \phi^p_{i_1 i_2} \end{aligned} \tag{13}$$

An estimate of ϕ^f is then obtained for every pair i_1, i_2 as

$$\hat{\phi}^f_{i_1,i_2} = \frac{(\hat{\phi}_{i_1 i_2} - \phi^p_{i_1 i_2})}{\kappa_{i_1 i_2}} \tag{14}$$

These pair-specific estimates can then be averaged over all pairs of sequenced subjects from the same population to obtain a global $\hat{\phi}^f$.

This second method of approximation relates the estimated mean kinship $\hat{\phi}^f$ to the distribution of the number of alleles distinct by descent in the founders. Then, $P[F_j, F_k]$ and $P[F_j]$ are derived from that distribution. The rest of this sub-subsection explains in detail how to compute the approximate values of these quantities.

The number of alleles A distinct by descent in the founders can take values $1, \ldots, 2n_f$. We will assume only the values $2n_f - 2, 2n_f - 1$ and $2n_f$ have nonzero probability. We parameterize the probabilities $P[A]$ to be proportional to

$$\begin{array}{ccc} 2n_f - 2 & 2n_f - 1 & 2n_f \\ \frac{1}{2}\theta^2 & \theta & 1 \end{array} \tag{15}$$

inspired from a truncated Poisson distribution. The expected kinship coefficient among the n_f founders is then

$$E[\Phi] = \frac{\theta \bar{\phi}_{2n_f-1} + \frac{1}{2}\theta^2 \bar{\phi}_{2n_f-2}}{1 + \theta + \frac{1}{2}\theta^2} \tag{16}$$

where $\bar{\phi}_a$ is the mean kinship coefficient among the n_f founders when there are a alleles distinct by descent. Assuming no inbreeding among the founders:

$$\begin{aligned}
\bar{\phi}_a &= \frac{1}{2}P[\text{Any founder shares an allele IBD with 2 other founders}] \\
&\quad +\frac{1}{4}P[\text{Any founder shares an allele IBD with 1 other founder}] \\
&= \frac{1}{2}\frac{2n_f - a}{n_f}\frac{2n_f - a - 1}{n_f - 1} \\
&\quad +\frac{1}{4}\frac{(a-n_f)(4n_f-2n_f)}{(a-n_f)(4n_f-2n_f)+(a-n_f)(2(a-n_f)-1)+(4n_f-2n_f)(2n_f-a-1)}
\end{aligned} \tag{17}$$

Equating $E[\Phi] = \hat{\phi}^f$, we solve for θ:

$$\hat{\theta} = \frac{-(\hat{\phi}^f - \bar{\phi}_{2n_f-1}) - \sqrt{(\hat{\phi}^f - \bar{\phi}_{2n_f-1})^2 - 2(\hat{\phi}^f - \bar{\phi}_{2n_f-2})\hat{\phi}^f}}{\hat{\phi}^f - \bar{\phi}_{2n_f-2}} \tag{18}$$

We then need $P[F_j, F_k|A = a]$:

$$\begin{aligned}
P[F_j, F_k|A = a] &= P[\text{Allele shared is RV}|j\&k \text{ share allele IBD}, A = a]P[j\&k \text{ share allele IBD}|A = a] \\
&= \frac{1}{a}2\bar{\phi}_a
\end{aligned} \tag{19}$$

Finally, $P[F_j, F_k]$ is obtained as

$$\begin{aligned}
P[F_j, F_k] &= \sum_{a=2nf-2}^{2nf} P[F_j, F_k|A = a]P[A = a] \\
&= \left(\frac{2\bar{\phi}_{2n_f-2}\frac{1}{2}\theta^2}{2n_f - 2} + \frac{2\bar{\phi}_{2n_f-1}\theta}{2n_f - 1}\right)\frac{1}{1+\theta+\frac{1}{2}\theta^2} \\
&= \left(\frac{\bar{\phi}_{2n_f-2}\theta^2}{2n_f - 2} + \frac{2\bar{\phi}_{2n_f-1}\theta}{2n_f - 1}\right)\frac{1}{1+\theta+\frac{1}{2}\theta^2}
\end{aligned} \tag{20}$$

The marginal probability that any founder j introduces the RV is

$$\begin{aligned} P[F_j] &= \sum_{a=2nf-2}^{2nf} P[F_j|A=a]P[A=a] \\ &= \left(\frac{\frac{1}{2}\theta^2}{2n_f-2} + \frac{\theta}{2n_f-1} + \frac{1}{2n_f}\right)\frac{1}{1+\theta+\frac{1}{2}\theta^2} \end{aligned} \tag{21}$$

As with method 1, the probability of F_j^U, the event that founder j is the only one to introduce the RV in the family, is obtained using equation 12.

When $\hat{\phi}^f$ is high, the approximation could be improved by allowing up to three alleles to be shared, i.e. give the event $A = 2n_f - 3$ a nonzero probability equal to $\frac{1}{6}\theta^3$. The estimation of θ would then require to solve a cubic polynomial, but the other formulas generalize easily.

1.2.2. Sharing probabilities conditional on the introduction of the RV by two of the founders

We need to introduce an additional type of subjects, the descendants that are common to the two founders introducing the RV, and who can therefore receive two copies of the variant. We note the number of copies of the RV in such a subject h by T_h.

As before, we begin by the expressions for the special case where all the sequenced subjects descend from every founder among their ancestors through independent lines of descent. With two founders introducing the RV, we further need to distinguish four events.

The lines of descent to every sequenced subject are common to the two founders introducing the variant

This implies that the two founders introducing the RV are mates and their descendants in common are their children. With the assumption of independent lines of descent, the n sequenced individuals descend from n children of the founders and

$$\begin{aligned} P[C_1 = \ldots = C_n = 1|F_j, F_k] &= \sum_{x=0}^{n} P[C_1 = \ldots = C_n = 1|\sharp\{i : T_i = 2\} = x, \sharp\{i : T_i = 1\} = n - x, F_j, F_k] \\ &\quad P[\sharp\{i : T_i = 2\} = x, \sharp\{i : T_i = 1\} = n - x|F_j, F_k] \\ &= \sum_{x=0}^{n} \left(\frac{1}{2}\right)^{\sum_{\{i:T_i=2\}} D_{ij}-2} \left(\frac{1}{2}\right)^{\sum_{\{i:T_i=1\}} D_{ij}-1} \binom{n}{x} \left(\frac{1}{4}\right)^{x} \left(\frac{1}{2}\right)^{n-x} \\ &= \sum_{x=0}^{n} \left(\frac{1}{2}\right)^{D^s-n-x} \binom{n}{x} \left(\frac{1}{2}\right)^{2x} \left(\frac{1}{2}\right)^{n-x} = \left(\frac{1}{2}\right)^{D^s} \sum_{x=0}^{n} \binom{n}{x} \\ &= \left(\frac{1}{2}\right)^{D^s-n} \end{aligned} \tag{22}$$

where $D^s = \sum_i D_{ij}$ and $D_{ij} = D_{ik} \forall i$. This expression applies if all $D_{ij} \geq 2$, i.e. the sequenced subjects are grand-children or more distant descendants of the founders. When a sequenced subject is a children of the founders, then $C_i = T_i$. We adapt the formula to distinguish the n_c sequenced subjects who are children of the founders from the others.

$$\begin{aligned} P[C_1 \geq 1, \ldots, C_{n_c} \geq 1, C_{n_c+1} = \ldots = C_n = 1|F_j, F_k] &= P[C_1 \geq 1, \ldots, C_{n_c} \geq 1|F_j, F_k] \\ &\quad P[C_{n_c+1} = \ldots = C_n = 1|F_j, F_k] \\ &= \left(\frac{3}{4}\right)^{n_c} \left(\frac{1}{2}\right)^{(D^s-n_c)-(n-n_c)} \\ &= \left(\frac{3}{4}\right)^{n_c} \left(\frac{1}{2}\right)^{D^s-n} \end{aligned} \tag{23}$$

The expression for the probability of not seeing the variant in any sequenced individual

when all $D_{ij} \geq 2$ is:

$$\begin{aligned} P[C_1 = \ldots = C_n = 0 | F_j, F_k] &= \sum_{x=0}^{n} \sum_{y=0}^{n-x} P[C_1 = \ldots = C_n = 0 | \sharp\{i : T_i = 2\} = x, \sharp\{i : T_i = 1\} = y, F_j, F_k] \\ & \quad P[\sharp\{i : T_i = 2\} = x, \sharp\{i : T_i = 1\} = y | F_j, F_k] \\ &= \sum_{x=0}^{n} \sum_{y=0}^{n-x} \prod_{\{i:T_i=2\}} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}-2}\right) \prod_{\{i:T_i=1\}} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}-1}\right) \\ & \quad \binom{n}{x, y, n-x-y} \left(\frac{1}{4}\right)^{x} \left(\frac{1}{2}\right)^{y} \left(\frac{1}{4}\right)^{n-x-y} \end{aligned} \tag{24}$$

without obvious simplification. The modification for sequenced subjects who are children of the founders is similar to that for the joint sharing probability, with probability equal to $\frac{1}{4}$ of not receiving the variant instead of $\frac{3}{4}$ of receiving it.

One founder is ancestor of all sequenced subjects and the other is ancestor of only one subject

We note j the founder who is ancestor of all sequenced subjects and 1 the sequenced subject descendant of the two founders j and k. There is only one child of founder k who can receive two copies of the variant (possibly subject 1 himself) and we note that child h. The number of copies he received is noted T.

$$\begin{aligned} P[C_1 = \ldots = C_n = 1 | F_j, F_k] &= P[C_1 = \ldots = C_n = 1 | T = 2, F_j, F_k] P[T = 2 | F_j, F_k] \\ & \quad + P[C_1 = \ldots = C_n = 1 | T = 1, F_j, F_k] P[T = 1 | F_j, F_k] \\ &= \left(\frac{1}{2}\right)^{D_{1h}-1+\sum_{i=2}^{n} D_{ij}} \left(\frac{1}{2}\right)^{D_{hj}} \frac{1}{2} \\ & \quad + \left(\frac{1}{2}\right)^{D_{1h}+\sum_{i=2}^{n} D_{ij}} \left[\left(\frac{1}{2}\right)^{D_{hj}} \frac{1}{2} + \left(1 - \left(\frac{1}{2}\right)^{D_{hj}}\right) \frac{1}{2}\right] \end{aligned} \tag{25}$$

$$= \left(\frac{1}{2}\right)^{D_{1h}+\sum_{i=2}^{n} D_{ij}} \left[\left(\frac{1}{2}\right)^{D_{hj}} + \frac{1}{2}\right]$$

This expression applies if $D_{1h} \geq 1$, i.e. subject 1 is not h himself, he or she is a grand-child or more distant descendant of the founder k. When subject 1 is a child of founder k, the expression becomes:

$$\begin{aligned} P[C_1 \geq 1, C_2 = \ldots = C_n = 1|F_j, F_k] &= P[C_1 = 2, C_2 = \ldots = C_n = 1|F_j, F_k] \\ &\quad + P[C_1 = \ldots = C_n = 1|F_j, F_k] \\ &= \left(\frac{1}{2}\right)^{D^s} \frac{1}{2} + \left(\frac{1}{2}\right)^{\sum_{i=2}^{n} D_{ij}} \frac{1}{2} \\ &= \left(\frac{1}{2}\right)^{D^s+1} \left[1 + 2^{D_{1j}}\right] \end{aligned} \tag{26}$$

The expression for the probability of not seeing the variant in any sequenced subject when $D_{ih} \geq 1$ is:

$$\begin{aligned} P[C_1 = \ldots = C_n = 0|F_j, F_k] &= \prod_{i=1}^{n} P[C_i = 0|F_j, F_k] \\ &= \begin{bmatrix} P[C_1 = 0|T = 2, F_j, F_k]P[T = 2|F_j, F_k] \\ +P[C_1 = 0|T = 1, F_j, F_k]P[T = 1|F_j, F_k] \\ +P[C_1 = 0|T = 0, F_j, F_k]P[T = 0|F_j, F_k] \end{bmatrix} \prod_{i=2}^{n} P[C_i = 0|F_j] \\ &= \begin{bmatrix} \left(1 - \left(\frac{1}{2}\right)^{D_{1h}-1}\right)\left(\frac{1}{2}\right)^{D_{hj}} \frac{1}{2} \\ \quad + \left(1 - \left(\frac{1}{2}\right)^{D_{1h}}\right) \frac{1}{2} \\ \quad + \left(1 - \left(\frac{1}{2}\right)^{D_{hj}}\right) \frac{1}{2} \end{bmatrix} \prod_{i=2}^{n} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}}\right) \end{aligned} \tag{27}$$

The same probability when subject 1 is a child of founder k is

$$P[C_1 = \ldots = C_n = 0|F_j, F_k] = \prod_{i=1}^{n} P[C_i = 0|F_j, F_k] = \frac{1}{2} \prod_{i=1}^{n} \left(1 - \left(\frac{1}{2}\right)^{D_{ij}}\right) \tag{28}$$

Each founder is ancestor of his own independent sequenced subject

We assume that founder j is ancestor of subject 1 and founder k is ancestor of subject 2. If there are $n > 2$ sequenced subjects, then $P[C_1 = \ldots = C_n = 1|F_j, F_k] = 0$. If $n = 2$, then

$$P[C_1 = C_2 = 1|F_j, F_k] = P[C_1 = 1|F_j]P[C_2 = 1|F_k] = \left(\frac{1}{2}\right)^{D_{1j}+D_{2k}} \tag{29}$$

The expression for the probability of not seeing the variant in any sequenced subject is

$$\begin{aligned} P[C_1 = \ldots = C_n = 0|F_j, F_k] &= P[C_1 = 0|F_j]P[C_2 = 0|F_k] \\ &= \left(1 - \left(\frac{1}{2}\right)^{D_{1j}}\right)\left(1 - \left(\frac{1}{2}\right)^{D_{2k}}\right) \end{aligned} \tag{30}$$

1.2.3. Obtaining an estimated sharing probability

We get an adjusted estimate of sharing probability with the following formula:

$$P[\text{RV shared}] = \frac{\sum_{j=1}^{n_f} P[C_1 = \ldots = C_n = 1|F_j^U]P[F_j^U] + \sum_j \sum_{k>j} P[C_1 = \ldots = C_n = 1|F_j, F_k]P[F_j, F_k]}{\sum_{j=1}^{n_f} P[C_1 + \ldots + C_n \geq 1|F_j^U]P[F_j^U] + \sum_j \sum_{k>j} P[C_1 + \ldots + C_n \geq 1|F_j, F_k]P[F_j, F_k]} \tag{31}$$