Relatedness and the X chromosome

David Wakeham

Walter and Eliza Hall Institute
Parkville, Victoria 3000

wakeham@wehi.edu.au

March 15, 2012

Introduction

I'll be discussing my work as a UROP student over past seven months.

Summary:

1. IdCoefs and ibd_cli
2. Extending Lange's algorithm to X chromosome
3. Gene dropping simulator
4. Continuous-time HMM for relatedness on X chromosome

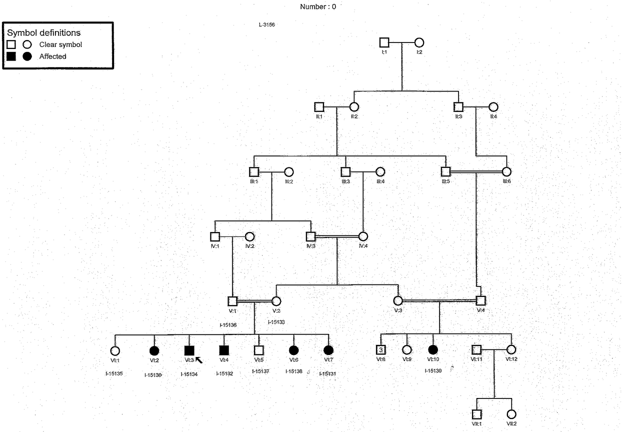
The recurring theme is relatedness, in particular on X chromosome.

IdCoefs and ibd_cli

My first project was finding pairwise (autosomal) IBD coefficients ω_k, $1 \leq k \leq 3$, given individuals and pedigree. This is likelihood two individuals will have k alleles IBD at a randomly chosen locus, conditional on common ancestry.

This information can be used as a QC measure, e.g., with PLINK, to detect misspecification of pedigrees or bad data. PLINK only estimates autosomal IBD vector from data.

You can calculate these numbers by hand for simple first- or second-order relations. But what about large, complicated pedigrees with inbreeding loops?


Number : 0
L-3156
Symbol definitions
□ ○ Clear symbol
■ ● Affected
I:1
I:2
II:1
II:2
II:3
II:4
III:1
III:2
III:3
III:4
III:5
III:6
IV:1
IV:2
IV:3
IV:4
V:1
V:2
V:3
V:4
I-15136
I-15133
VI:1
VI:2
VI:3
VI:4
VI:5
VI:6
VI:7
3
VI:8
VI:9
VI:10
VI:11
VI:12
I-15135
I-15130
I-15134
I-15132
I-15137
I-15138
I-15131
I-15139
VII:1
VII:2

IBD coefficients can be calculated recursively using algorithm in Kenneth Lange's *Mathematical and Statistical Methods for Genetic Analysis*. My initial plan was to code it from scratch.

Thankfully, Mark Abney (UChicago) had already done heavy lifting. He wrote fast program called `IdCoefs` for calculating *identity coefficients*, Δ_k, $1 \leq k \leq 9$. Autosomes only.

These are probabilities of specific allele configurations occurring, conditional on relationship. I'll go into more detail about these configurations in the next section of the talk. IBD coefficients can be expressed as sums of identity coefficients.

In fact,

$$\begin{aligned}\omega_0 &= \Delta_2 + \Delta_4 + \Delta_6 + \Delta_9 \\ \omega_1 &= \Delta_3 + \Delta_5 + \Delta_8 \\ \omega_2 &= \Delta_1 + \Delta_7.\end{aligned}$$

Kinship coefficients are also (indirectly) function of Δ_k:

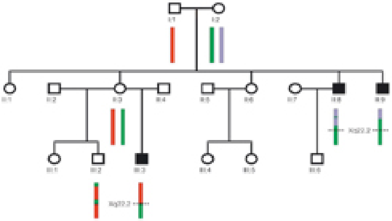
$$\Phi_{ij} = \Delta_1 + \tfrac{1}{2}\Delta_3 + \tfrac{1}{2}\Delta_5 + \tfrac{1}{2}\Delta_7 + \tfrac{1}{4}\Delta_8,$$

where Δ_1^m is the inbreeding coefficient for m.

I wrote a command line interface `ibd_cli` in Python to talk to `IdCoefs`. It takes ped file, feeds it to `IdCoefs`, processes the output and returns IBD coefficients. So it's basically a hack. Keith recently installed `IdCoefs` on unices (makes life much easier) and code for `ibd_cli` is maintained with SVN.

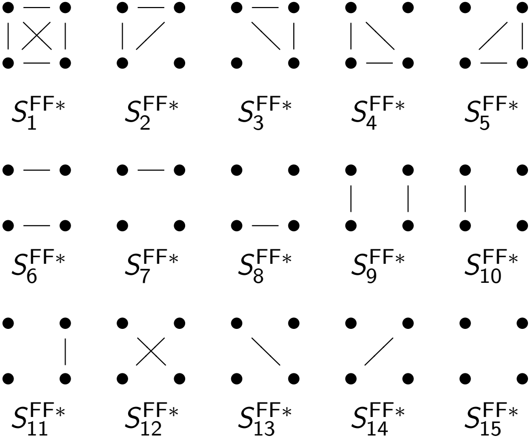
Xtending Lange's algorithm

What if we want to calculate IBD coefficients for X chromosome? Melanie discussed some practical applications in lab talk, December last year. E.g., Keipert syndrome X-linked, want IBD coefficients for pedigree:



In fact, extending algorithm to X case is one of the exercises for Chapter 5 of Lange.

We can extend Lange's approach in straightforward way. Start off by defining *identity states* S_k^* for cases FF, FM, and MM. These are specific allele configurations. First, 15 FF IDs $S_k^{\mathrm{FF}*}$,

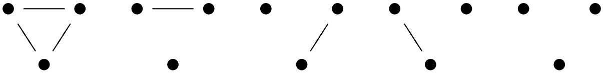


$S_1^{\mathrm{FF}*}$ $S_2^{\mathrm{FF}*}$ $S_3^{\mathrm{FF}*}$ $S_4^{\mathrm{FF}*}$ $S_5^{\mathrm{FF}*}$

$S_6^{\mathrm{FF}*}$ $S_7^{\mathrm{FF}*}$ $S_8^{\mathrm{FF}*}$ $S_9^{\mathrm{FF}*}$ $S_{10}^{\mathrm{FF}*}$

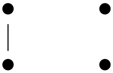
$S_{11}^{\mathrm{FF}*}$ $S_{12}^{\mathrm{FF}*}$ $S_{13}^{\mathrm{FF}*}$ $S_{14}^{\mathrm{FF}*}$ $S_{15}^{\mathrm{FF}*}$

Now the 5 FM IDs $S_k^{\text{FM}*}$,



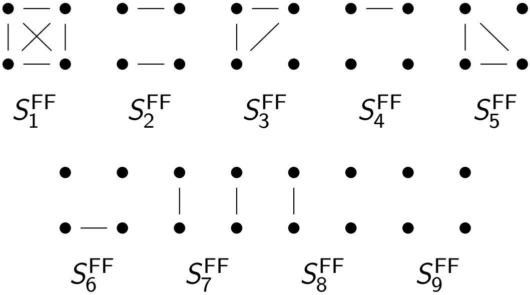
$S_1^{\text{FM}*}$ $S_2^{\text{FM}*}$ $S_3^{\text{FM}*}$ $S_4^{\text{FM}*}$ $S_5^{\text{FM}*}$

Finally, 2 MM IDs $S_k^{\text{MM}*}$,

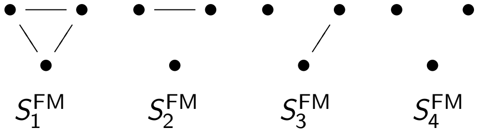


$S_1^{\text{MM}*}$ $S_2^{\text{MM}*}$

We then collapse these identity states into equivalence classes modulo reordering of genotypes. These are *condensed identity states* S_k. Again, start with 9 FF CIDs, S_k^{FF},



There are 4 FM CIDs S_k^{FM}



For the MM case, $S_k^{\text{MM}} = S_k^{\text{MM}*}$. The identity coefficients Δ_k are defined by

$$\Delta_k = P(S_k \mid \text{common ancestry}).$$

Simple relation to IBD coefficients. For FF case, relations given on slide 6. For FM case,

$$(\omega_0, \omega_1, \omega_2) = (\Delta_2^{\text{FM}} + \Delta_4^{\text{FM}}, \Delta_1^{\text{FM}} + \Delta_3^{\text{FM}}, 0).$$

Finally, for MM case,

$$(\omega_0, \omega_1, \omega_2) = (\Delta_2^{\mathrm{MM}}, \Delta_1^{\mathrm{MM}}, 0).$$

So, to calculate IBD coefficients we need identity coefficients. However, as with IBD coefficients, can't get directly. We need some more functions!

Let Ψ_k be probability condensed identity state S_k describes random sample *with replacement* of c_i alleles for both individuals at (randomly chosen) locus on X, where c_i is no. of copies of X individual i has. This is not the same as Δ_k, since sampling occurs without replacement in that case.

We write vectors of Ψ_k and Δ_k as $\vec{\Psi}$, $\vec{\Delta}$ resp. Using basic probability theory, we can write $\vec{\Psi}$ in terms of the $\vec{\Delta}$. In matrix form, for FF case, we get

$$\vec{\Psi} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{8} & \frac{1}{16} & 0 \\ 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{16} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{8} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{8} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix} \vec{\Delta}.$$

The coefficient matrix is invertible (upper triangular, determinant product of diagonal elements), so we can express $\vec{\Delta}$ in terms of $\vec{\Psi}$.

Inverting, we get

$$\vec{\Delta} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & -\frac{1}{2} & -1 & \frac{1}{2} & \frac{3}{4} & 1 \\ 0 & 0 & 2 & 0 & 0 & 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 2 & 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix} \vec{\Psi}.$$

Sanity check, column entries sum to 1 (not stochastic, though).

We can do same in FM and MM case. FM case:

$$\vec{\Psi} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \vec{\Delta} \quad \Longrightarrow \quad \vec{\Delta} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \vec{\Psi}.$$

For MM case, $\vec{\Psi} = \vec{\Delta}$. So, MM situation is, as you would expect, mathematically uninteresting.

Now we find a way to calculate $\vec{\Psi}$. We need what Lange calls *generalised kinship coefficients*. This is really just probability measure on set of sampled genes, and, as the name suggests, can be used for much more than pairwise comparison.

If P is partition of sampled genes, we define $\Phi(P)$ to be probability that IBD relation induces P on sampled genes. Let G_i^n denote the n-th allele sampled from individual i at locus on X. In fact, Ψ_k is always an integer multiple of $\Phi(P)$ for some P (multiplier is no. of states in condensed state).

For FF, i and j are female and we sample $G_i^1, G_i^2, G_j^1, G_j^2$.

$$\begin{aligned}
\Psi_1^{\text{FF}} &= \Phi(\{G_i^1, G_i^2, G_j^1, G_j^2\}) & \Psi_2^{\text{FF}} &= \Phi(\{G_i^1, G_i^2\}\{G_j^1, G_j^2\}) \\
\Psi_3^{\text{FF}} &= 2\Phi(\{G_i^1, G_i^2, G_j^1\}\{G_j^2\}) & \Psi_4^{\text{FF}} &= \Phi(\{G_i^1, G_i^2\}\{G_j^1\}\{G_j^2\}) \\
\Psi_5^{\text{FF}} &= 2\Phi(\{G_i^1, G_j^1, G_j^2\}\{G_i^2\}) & \Psi_6^{\text{FF}} &= \Phi(\{G_i^1\}\{G_i^2\}\{G_j^1, G_j^2\}) \\
\Psi_7^{\text{FF}} &= 2\Phi(\{G_i^1, G_j^1\}\{G_i^2, G_j^2\}) & \Psi_8^{\text{FF}} &= 4\Phi(\{G_i^1, G_j^1\}\{G_i^2\}\{G_j^2\}) \\
\Psi_9^{\text{FF}} &= \Phi(\{G_i^1\}\{G_j^1\}\{G_i^2\}\{G_j^2\}).
\end{aligned}$$

For FM, let i be female and j male. Now the sampled genes are G_i^1, G_i^2, G_j, and

$$\begin{array}{ll} \Psi_1^{\mathrm{FM}} = \Phi(\{G_i^1, G_i^2, G_j\}) & \Psi_2^{\mathrm{FM}} = \Phi(\{G_i^1, G_i^2\}\{G_j\}) \\ \Psi_3^{\mathrm{FM}} = 2\Phi(\{G_i^1, G_j\}\{G_i^2\}) & \Psi_4^{\mathrm{FM}} = \Phi(\{G_i^1\}\{G_i^2\}\{G_j\}). \end{array}$$

Finally, for MM, $\Psi_1^{\mathrm{MM}} = \Phi(\{G_i, G_j\})$ and $\Psi_2^{\mathrm{MM}} = \Phi(\{G_i\}\{G_j\})$.

We can find $\Phi(P)$ using boundary conditions and recurrence relations. Some terminology. A *block* is a member of a partition. M, F = male, female, used as adjective and noun. Finally, j is the mother and k the father of i.

Boundary Conditions

(B1) If F is involved in ≥ 3 or M is involved ≥ 2 blocks, $\Phi(P) = 0$. Can't have that many X genes pairwise non-IBD.

(B2) If genes sampled from distinct founders occur in same block, then $\Phi(P) = 0$. Founders aren't related!

(B3) If only founders contribute sampled genes, and neither (B1), (B2) apply, then

$$\Phi(P) = 2^{m_2 - m_1}$$

where m_1 is total number of genes sampled from F founders and m_2 is number of F founders sampled. There are $m_1 - m_2$ "comparison events", iid Bn(0.5).

Recurrence Rules

(R1) Assume $G_i^1, \ldots, G_i^s$ are sampled from i for $s \geq 1$ and occur in one block, $P = \{\{G_i^1, \ldots, G_i^s, \ldots\}\} \cup P'$. If i is M, then

$$\Phi(\{G_i^1, \ldots, G_i^s, \ldots\}, P') = \Phi(\{G_j, \ldots\}, P')$$

since i receives X gene from j. If i is F, then

$$\begin{aligned}\Phi(\{G_i^1, \ldots, G_i^s, \ldots\}, P') = (1 - 2^{1-s})&\Phi(\{G_j, G_k, \ldots\}, P') \\ + 2^{-s}&\Phi(\{G_j, \ldots\}, P') \\ + 2^{-s}&\Phi(\{G_k, \ldots\}, P')\end{aligned}$$

since there is 2^{-s} chance all sampled from j and 2^{-s} chance all sampled from k. (LoTP, condition on parental sampling.)

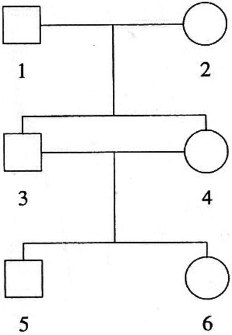
(R2) Now assume genes $G_i^1, \ldots, G_i^s, G_i^{s+1}, \ldots, G_i^{s+t}$ sampled from F i and $P = \{\{G_i^1, \ldots, G_i^s, \ldots\}, \{G_i^{s+1}, \ldots, G_i^{s+t}, \ldots\}\} \cup P'$. Then

$$\begin{aligned}\Phi(\{G_i^1, \ldots, G_i^s, \ldots\}\{G_i^{s+1}, \ldots, G_i^{s+t}, \ldots\}, P') \\ = 2^{-(s+t)}\Phi(\{G_j, \ldots\}\{G_k, \ldots\}, P') \\ + 2^{-(s+t)}\Phi(\{G_k, \ldots\}\{G_j, \ldots\}, P').\end{aligned}$$

This is because there is $2^{-(s+t)}$ chance that genes in first block all come from j, all in second from k. Similarly with j and k swapped. No male analogue due to (B1).

I've implemented whole shebang in Python, coming to SVN soon!

We'll do a sample calculation to give a feel for way formalism works. In practice, use computer. Consider inbred pedigree



Suppose we want to find Ψ_3 for 2 and 5.

(I chickened out from doing comparison between second and third generation since there is *no board*.)

We have

$$\begin{aligned}
\Psi_3 &= 2\Phi(\{G_2, G_5\}\{G_2\}) \\
&= 2\Phi(\{G_2, G_4\}\{G_2\}) && \text{(R1)} \\
&= \tfrac{1}{2}\Phi(\{G_2\}\{G_2\}) + \tfrac{1}{2}\Phi(\{G_2, G_1\}\{G_2\}) && \text{(R2)} \\
&= \tfrac{1}{2}\Phi(\{G_2\}\{G_2\}) && \text{(B2)} \\
&= \tfrac{1}{2} \cdot \tfrac{1}{2} = 0.25. && \text{(B3)}
\end{aligned}$$

This makes sense, since $1/2$ chance real configuration is S_3, and $1/2$ chance we pick different alleles from individual 2. We can also double check our calculation with program.

Gene dropper

To test extension of Lange's algorithm, I wrote Python script to drop genes (autosomes + X) on a pedigree.

Jim Stankovich has written program for autosomal case (+ whole battery of statistical tests), but in foreign languages. So, written from the ground up.

Basically, program treats chromosomes as unit intervals, crossovers as spatial Poisson process, etc. Ignores linkage, psudoautosomal crossovers. Chromosomes are partitioned, blocks tagged by founder. Shared content is calculated by checking labels and adding length of shared sections.

The empirical IBD values obtained using gene dropper agreed with values predicted by my algorithm. Great!

I also used the gene dropper to test idea Melanie had. We thought that, although relatedness signal on X for related individuals would be initially much weaker than an autosomal signal, it would decay more slowly due to MF meioses. In some cases, it might "outlast" autosomal signal, in the sense that normalised length of shared segments is larger. Question: Is this true? If so, when?

Answer: When a long path with *no* MM meioses and lots of MF meioses joins individuals. For instance, in a non-inbred pedigree, an "alternating path" MFMF... of length > 10 will usually do it. We can use this to look for pedigrees apposite to analysis.

HMM for relatedness on X

Albrechtsen (2009) introduced cleverly parameterised hidden Markov model (HMM) to perform pairwise relatedness mapping on autosomes. Lately, I've been working on extending model to X chromosome and implementing it.

Reminder: HMMs have a *hidden* Markov chain/process concealed from observer and *observation* process, where distribution depends on the hidden state at time of observation. Hidden process is stochastic IBD process, assumed to be continuous, time-homogeneous, and Markovian. In reality, not generally Markovian, but good approximation for inferential purposes. Observations are pairs of genotypes, i.e., looking at 2 – 4 alleles total.

The model can be used to estimate two things: *global relatedness* in the form of IBD coefficients, and *local relatedness*, posterior probability individuals have particular IBD state at a chosen locus.

Let's get a bit more formal. First, we assume that individuals are distantly related and not inbred, so $\omega_2 = 0$ and $\omega_0 + \omega_1 = 1$. The transition rate matrix for the IBD process is as follows:

$$\mathbf{Q} = \alpha \begin{bmatrix} -\omega_1 & \omega_1 \\ \omega_0 & -\omega_0 \end{bmatrix}.$$

We denote IBD states by *Zi*, or *i* if unambiguous. Entry q_{ij} of $\mathbf{Q}$ is rate of going from *Zi* to *Zj*. Stationary distribution is $\pi = (\omega_0, \omega_1)$. Matrix makes sense, rate of going from *Zi* to *Zj*, $i \neq j$, proportional to ω_j.

α is an unknown parameter which controls the frequency of transitions. If individuals are joined by n paths of ℓ meioses (and no other paths), outbred ancestors, then

$$\alpha = -\ell^{2-n} \log(1-\theta)$$

where θ is recombination rate. See Purcell (2007), Albrechtsen (2009) for calculations. Relation is not simple for more complicated pedigree graphs or X case.

Let $\mathbf{P}(t) = [p_{kl}(t)]$ be the *time-dependent transition matrix*, i.e., matrix of transition probabilities for a time interval t, where

$$p_{kl}(t) = P(X(t) = l | X(0) = k).$$

Sensible to set $\mathbf{P}(0) = \mathbf{I}$. We can find $\mathbf{P}(t)$ using the *Kolmogorov equations*,

$$\frac{\partial}{\partial t}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q} = \mathbf{Q}\mathbf{P}(t).$$

Assuming $\mathbf{Q}$ is diagonalisable, these can be solved to give

$$\mathbf{P}(t) = \exp(t\mathbf{Q}) = \mathbf{\Lambda} \exp(t\mathbf{D})\mathbf{\Lambda}^{-1} = \mathbf{\Lambda} \left(\sum_{n \geq 0} \frac{t^n}{n!} \mathbf{D}^n \right) \mathbf{\Lambda}^{-1}$$

where $\mathbf{D}$ is diagonal and $\mathbf{Q} = \mathbf{\Lambda D \Lambda}^{-1}$. In our case, $\mathbf{Q}$ can be diagonalised, with

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & -\alpha \end{bmatrix}, \ \mathbf{\Lambda} = \begin{bmatrix} 1 & \alpha\omega_1 \\ 1 & -\alpha\omega_0 \end{bmatrix}.$$

After some algebraic manipulation, we get

$$\mathbf{P}(t) = \begin{bmatrix} \omega_0 + \omega_1 e^{-\alpha t} & \omega_1(1 - e^{-\alpha t}) \\ \omega_0(1 - e^{-\alpha t}) & \omega_1 + \omega_0 e^{-\alpha t} \end{bmatrix}.$$

The IBD process is the same for X and autosomal cases.

Emission process

The emission process is different for X case. In fact, unsurprisingly, it depends on sexes of the individuals being compared, since number of copies of X determine observation probabilities. This will give rise to three observation processes and therefore three flavours of HMM.

We assume marker loci are biallelic and select reference allele at each marker. The *genotypic state* for an individual at a marker counts reference alleles at that locus. Thus, genotypic states are elements of $\Phi = \{0, 1, 2\}$. An observation at marker m is ordered pair of genotypic states, denoted G_m, with $G_m \in \Phi^2$.

Emission probabilities are probabilities of observing given genotypic pair conditional on hidden IBD state. Let A and a be the marker alleles, $p = p_A$ and $q = p_a$ their respective population frequencies, and A the reference allele.

When individuals are both female (FF), we have

	G	$Z0$	$Z1$	$Z2$
AA AA	$(2,2)$	p^4	p^3	p^2
AA aa	$(2,0)$	p^2q^2	0	0
AA Aa	$(2,1)$	$2p^3q$	p^2q	0
Aa Aa	$(1,1)$	$2p^2q^2$	$p^2q + q^2p$	pq

Trick for calculating — terms of $(p+q)^{\sum c_i - \text{ibd}}$.

Similarly, if individuals of opposite sex (FM), we have

	G	$Z0$	$Z1$
AA A	$(2,1)$	p^3	p^2
AA a	$(2,0)$	$p^2 q$	0
Aa A	$(1,1)$	$2p^2 q$	$2pq$

Finally, if both individuals are male (MM), then

	G	$Z0$	$Z1$
A A	$(1,1)$	p^2	p
A a	$(1,0)$	pq	0

Swap A and a to obtain other emission probabilities.

Likelihood

Likelihood calculation is routine. Suppose we have M markers. Let $\mathbf{X} = X_1 X_2 \cdots X_M$, where $X_m \in \{0, 1\}$ denotes IBD state at position of mth marker. Let $\Omega = \{\omega_0, \omega_1\}$. For observation sequence $\mathbf{G} = G_1 G_2 \cdots G_M$, we have

$$P(\mathbf{G}|\Omega, \alpha) = \sum_{\mathbf{x}} f(\mathbf{G}, \mathbf{X}) P(G_1|X_1 = x_1) P(X_1 = x_1|\Omega)$$

where

$$f(\mathbf{G}, \mathbf{X}) = \prod_{m=2}^{M} P(G_m|X_m = x_m) P(X_m = x_m|X_{m-1} = x_{m-1}, \Omega, \alpha)$$

and $\mathbf{x} = x_1 x_2 \cdots x_M$ ranges over $\{0, 1\}^M$, i.e., all IBD chains.

In practice, use forward-backward algorithm. For more details, see Rabiner's *A Tutorial on Hidden Markov Models* (1989).

Global relatedness

Given an observation sequence **G**, we estimate parameters by maximising likelihood, e.g., finding

$$\operatorname{argmax} \theta(S)$$

where $S = \Omega \cup \alpha$ and $\theta(S) = P(\mathbf{G} \mid S)$. Because the underlying process is not discrete (and intermarker distances vary), the usual parameter reestimation method needs to be modified. In fact, easier just to go straight to numerical optimisation techniques, e.g., L-BFGS. Currently trying to get this to work with scipy/numpy.

Local relatedness

Once we have estimated parameters, easy to find most likely hidden state at given marker m. This is the local relatedness estimation. We use forward and backward variables from forward-backward algorithm. Define *forward variable*

$$\alpha_m(k) = P(G_1 G_2 \cdots G_m, X_m = k \mid \Omega, \alpha)$$

and *backward variable*

$$\beta_m(k) = P(G_{m+1} G_{m+2} \cdots G_M \mid X_m = k, \Omega, \alpha).$$

These can be calculated recursively; again, see Rabiner.

Now, define posterior probability that hidden state at *m*-th marker is *k*, i.e.,

$$\gamma_m(k) = P(X_m = k \mid \mathbf{G}, \Omega, \alpha).$$

Bayes' theorem and law of total probability imply

$$\gamma_m(k) = \frac{\alpha_m(k)\beta_m(k)}{\alpha_m(0)\beta_m(0) + \alpha_m(1)\beta_m(1)}.$$

Since it's easy to calculate α and β, this is also easy to calculate. Most likely state is then

$$\mathrm{argmax}_{k \in \{0,1\}}[\gamma_m(k)].$$

I'll be testing model on simulated data (script to generate marker data on given pedigree) and real data (e.g., NGS data on Keipert syndrome, Xomes from Josef Gécz) soon, hopefully.

Next step is probably doing case where inbreeding is allowed. Following Moltke (2011), this will involve MCMC methods, should be fun.

That's all folks. Questions?