

Parameters and pedigrees. Grand goal: Reconstruct family (pedigree) from DNA.

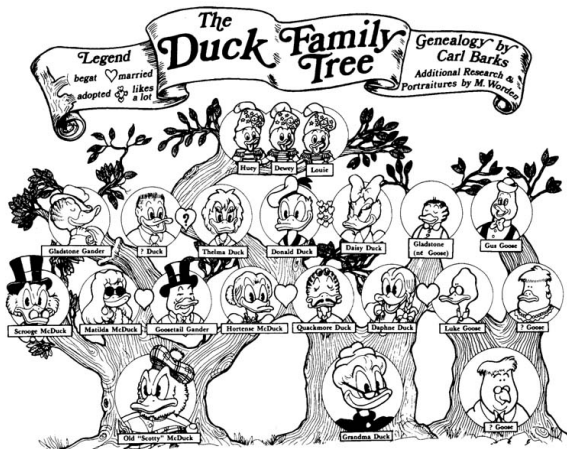
Thore Egeland^{(1),(2)}

(1) Norwegian University of Life Sciences, (2) OUS

BBB May 22-24, Oslo

Preface: What does it mean to be related?
Estimating relationships. Non-inbred, pairwise case
General pairwise case: Consanguinity, Jacquard
Adoption Case

Non-genetical approach



Preface: What does it mean to be related?

Estimating relationships. Non-inbred, pairwise case

General pairwise case: Consanguinity, Jacquard

Adoption Case

Genetical approach



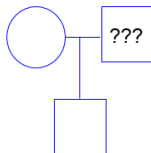
Preface: What does it mean to be related?

Estimating relationships. Non-inbred, pairwise case

General pairwise case: Consanguinity, Jacquard

Adoption Case

Applications: “Red Wine, Paternity and Disasters”



A
T
C
G

Common: Problem, data and statistical evaluation



More Applications

Statistical Science

2016, Vol. 31, No. 2, 259–274

DOI: 10.1214/16-STS552

In the Public Domain

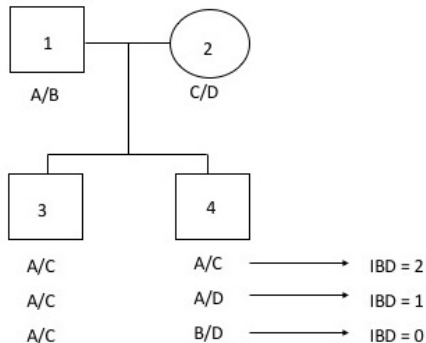
Close-Kin Mark-Recapture

Mark V. Bravington, Hans J. Skaug and Eric C. Anderson

Abstract. Mark-recapture (MR) methods are commonly used to study wildlife populations. Taking advantage of modern genetics one can generalize from “recapture of self” to “recapture of closely-related kin”. Abundance

Outline

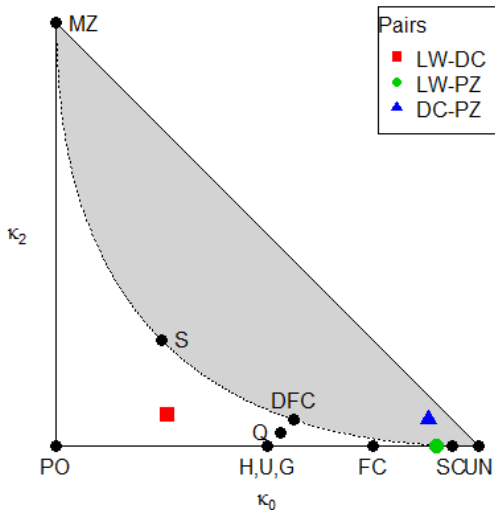
- ▶ **IBD**. Previous work [Thompson, 1975, Thompson, 2000].
- ▶ **Jacquard coefficients**:
Modelling consanguinity (inbreeding), [Jacquard, 1972].
- ▶ Examples: Simulation, **adoption case**.
- ▶ Beyond the pairwise case.



Non-inbred pairwise relationships are characterised by

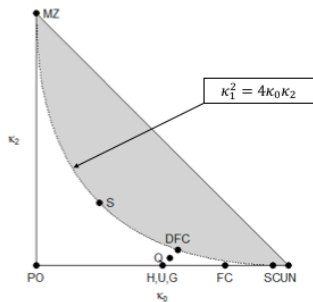
$$\kappa_0 = P(\text{IBD} = 0), \quad \kappa_2 = P(\text{IBD} = 2)$$

Adoption example - more later

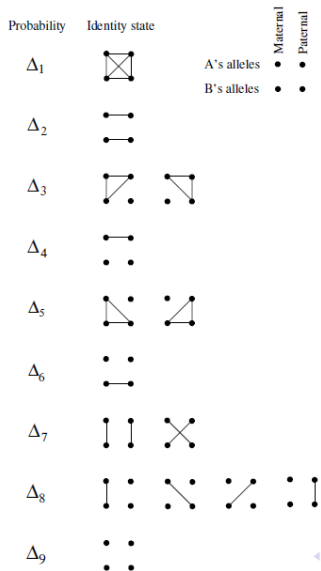


IBD triangle

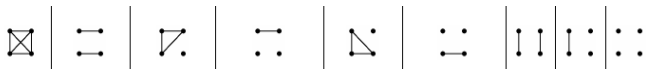
Relationship	$\kappa = (\kappa_0, \kappa_1, \kappa_2)$
Parent-child (PO)	(0, 1, 0)
Siblings (S)	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$
Avuncular (H, G, U)	$(\frac{1}{2}, \frac{1}{2}, 0)$
First cousins (FC)	$(\frac{3}{4}, \frac{1}{4}, 0)$
Double first cousins (DFC)	$(\frac{9}{16}, \frac{6}{16}, \frac{1}{16})$
Quadruple half first cousins (Q)	$(\frac{17}{32}, \frac{14}{32}, \frac{1}{32})$
Second cousins (SC)	$(\frac{15}{16}, \frac{1}{16}, 0)$
Unrelated (UN)	(1, 0, 0)
Monozygotic twins (MZ)	(0, 0, 1)



Boundary & corner asymptotics [García-Magariños et al., 2015].
 Challenges remain: [Kaur et al., 2017].



Probabilities given Jacquard State



$$\mathcal{P} = \begin{pmatrix}
 aa, aa & p & p^2 & p^2 & p^3 & p^2 & p^3 & p^2 & p^3 & p^4 \\
 aa, bb & 0 & pq & 0 & pq^2 & 0 & p^2q & 0 & 0 & p^2q^2 \\
 aa, ab & 0 & 0 & pq & 2p^2q & 0 & 0 & 0 & p^2q & 2p^3q \\
 aa, bc & 0 & 0 & 0 & 2pqr & 0 & 0 & 0 & 0 & 2p^2qr \\
 ab, aa & 0 & 0 & 0 & 0 & pq & 2p^2q & 0 & p^2q & 2p^3q \\
 bc, aa & 0 & 0 & 0 & 0 & 0 & 2pqr & 0 & 0 & 2p^2qr \\
 ab, ab & 0 & 0 & 0 & 0 & 0 & 0 & 2pq & pq(p+q) & 4p^2q^2 \\
 ab, ac & 0 & 0 & 0 & 0 & 0 & 0 & 0 & pqr & 4p^2qr \\
 ab, cd & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4pqrs
 \end{pmatrix}$$

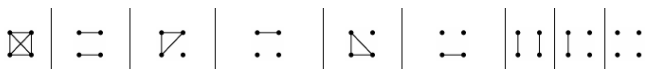
Likelihood. N independent markers

For each marker we observe $j \in \{1, \dots, 9\}$.

$$\text{Likelihood: } L(\Delta) = \prod_{i=1}^N \mathcal{P}(j_i,)\Delta$$

- ▶ This gives explicit expressions for log likelihood, gradient and Hessian.
- ▶ R library `maxLik` used for constrained MLE, `paramlink` (Vigeland) for simulation, plotting etc.
- ▶ **Problem for asymptotics and bootstrapping:** permissible parameter region not known, boundary issues.

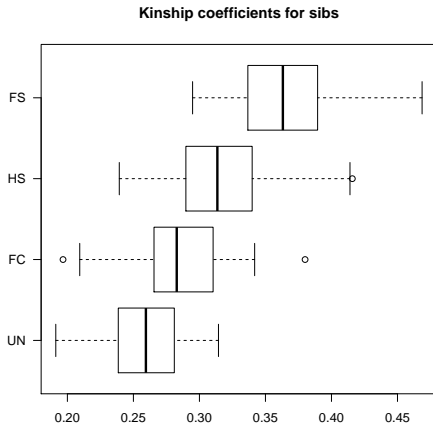
Kinship coefficient



$$\psi = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$$

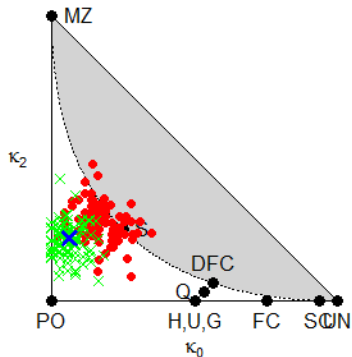
- ▶ **Interpretation:** Probability that random allele in one individual is IBD to random allele in another individual for a specific locus.
- ▶ **Equivalently:** Inbreeding coefficient of their child.

Example. Simulation: 50 markers each with 20 equifrequent alleles.



Parents: UN 0.25, FC: 0.28125, HS: 0.3125, FS: 0.375

Effect of Consanguinity



- ▶ Sibs simulated, parents sibs. Blue cross.
- ▶ Red points: estimates ignoring consanguinity.

Background. Current Data

Background: PZ was born October 1981 in Indonesia and soon thereafter adopted to the Netherlands. She is now searching for biological relatives. Two alleged sisters (not true) LW and DC have been genotyped for 33 markers.

	loc1.1	loc1.2	loc2.1	loc2.2	...	loc33.1	loc33.2
LW	8	9	3	3	...	7	14
DC	8	8	3	4	...	8	10
PZ	6	8	6	6	...	12	12

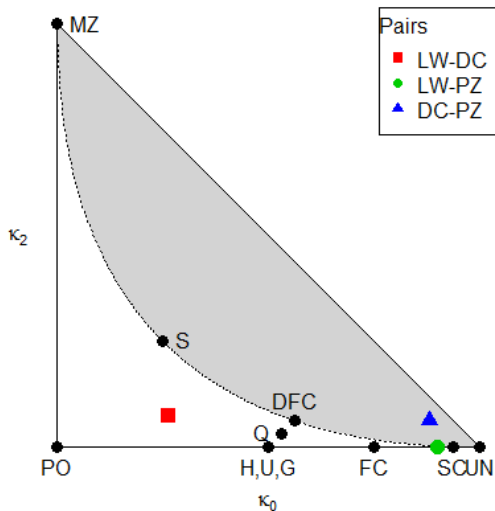
	1	2	...	11	12	...	21
loc1	0.0001	0.0006	...	0.0060	0.0008		
...							
loc33	0.0493	0.0006	...	0.0941	0.0778	...	0.0012

I: Familias approach (familias.name)

- ▶ Generate a few thousand pedigrees linking the three women:
 x_1, \dots, x_S .
- ▶ Assign priors $\pi(x_1), \dots, \pi(x_S)$.
Priors can accommodate inbreeding, promiscuity, ...
- ▶ Calculate and sort likelihoods (Elston-Stewart algorithm)
 $L_1 \leq \dots \leq L_S$.

Unfortunately, no pedigrees stick out

II: Ignoring Consanguinity



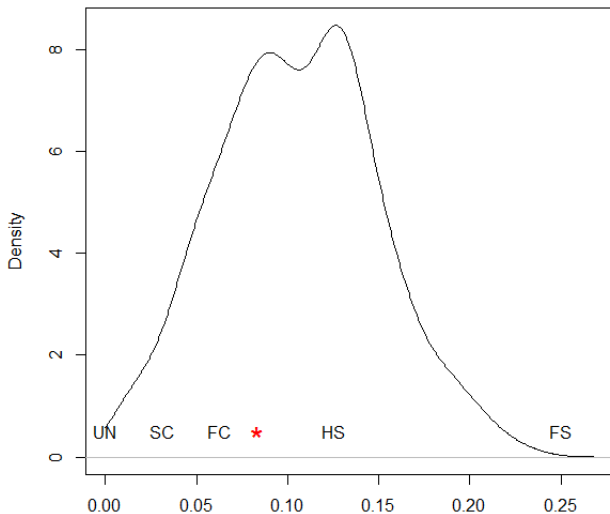
III: Kinship coefficient

Less ambitious goal: Consider kinship coefficient.

	Consanguinity		Ignoring Consanguinity	
	LW	DC	LW	DC
DC	0.212		0.202	
PZ	0.036	0.083	0.025	0.046

Table 1: Kinship coefficients. For reference: Second cousins: 0.03125, first cousins: 0.0625, half sibs: 0.125, full sibs: 0.25

Kinship estimates PZ-DC. 400 bootstrap simulations



- ▶ We could test if the kinship coefficient exceeds the background value (??) ψ_0 :
 - $H_0 : \psi \leq \psi_0$
 - $H_1 : \psi > \psi_0$
- ▶ Waiting for data ...
- ▶ More potential relatives, more markers will be available soon.

198,091 parameters needed for six relatives

total (N_n)	number	
	distinct (D_n)	D_n/N_n
2	2	1.0000
15	9	0.6000
203	66	0.3251
4140	712	0.1720
115,975	10,457	0.0902
4,213,597	<u>198,091</u>	0.0470



Egeland, T., Kling, D., and Mostad, P. (2015).

Relationship Inference with Familias and R: Statistical Methods in Forensic Genetics.
Academic Press.



Egeland, T., Pinto, N., and Vigeland, M. (2013).

A general approach to power calculation for relationship testing.
Forensic Science International: Genetics.



García-Magariños, M., Egeland, T., López-de Ullibarri, I., Hjort, N. L., and Salas, A. (2015).

A parametric approach to kinship hypothesis testing using identity-by-descent parameters.
Statistical Applications in Genetics and Molecular Biology, 14(5):465–479.



Jacquard, A. (1972).

Genetic information given by a relative.
Biometrics, pages 1101–1114.



Kaur, N., Storvik, G., Vigeland, M. D., and Egeland, T. (2017).

Relationship inference: Estimation and model selection.
Manuscript.



Thompson, E. (1986).

Pedigree Analysis in Human Genetics.
The Johns Hopkins University Press, Baltimore.



Thompson, E. (2000).

Statistical inference from genetic data on pedigrees.
In NSF-CBMS regional conference series in probability and statistics.



Thompson, E. A. (1975).

The estimation of pairwise relationships.
Annals of Human Genetics, 39:173–188.