Linkage in Mx & Merlin

Meike Bartels
Kate Morley
Hermine Maes

Based on Posthuma et al., Boulder & Egmond

Outline

- Summary of yesterday afternoon
- Linkage in Merlin – Phi-Hat
- Linkage in Mx – Mixture

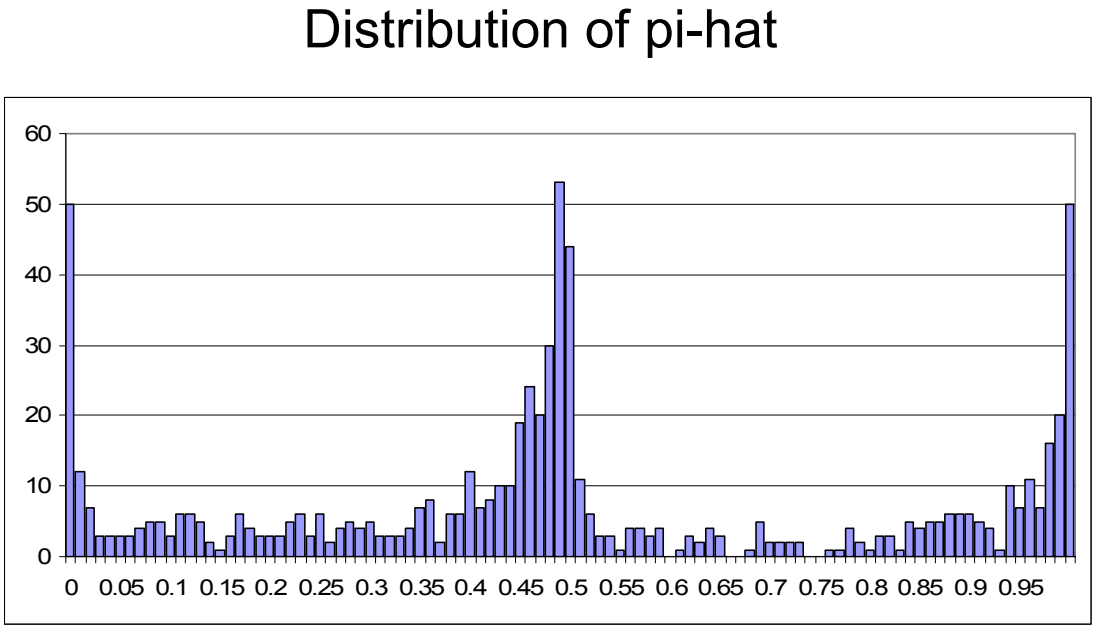
Summary Yesterday -Linkage Analysis

- Where are the genes?
  - Collect genotypic data on large number of markers
  - Compare correlations by number of alleles identical by descent at a particular marker
  - Partition/ Quantify variance in genetic (QTL) and environmental components
  - Test significance of QTL effect

Summary Yesterday - Methods

- Partitioned Twin Analyses
- Linkage using Pi-Hat

Partioned Twin Analyses



- DZ pairs: distribution of pi-hat (π) at particular cM on chromosome 2
- π<0.25: IBD=0 group π>0.75: IBD=2 group
  others: IBD=1 group
- picat= (0,1,2)

Partioned Twin Analyses

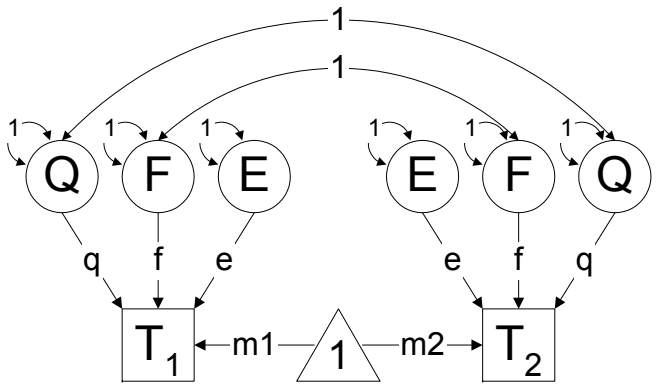
- Can resemblance (e.g. correlations, covariances) between sib pairs, or DZ twins, be modeled as a function of DNA marker sharing (IBD) at a particular chromosomal location?
  - Estimate covariance by IBD state
  - Impose genetic model and estimate model parameters

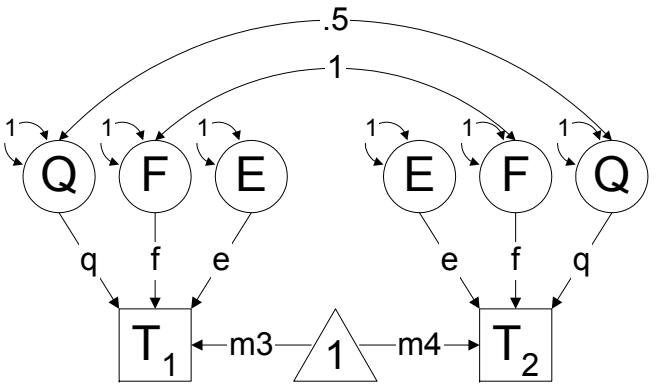
Correlations – model fit

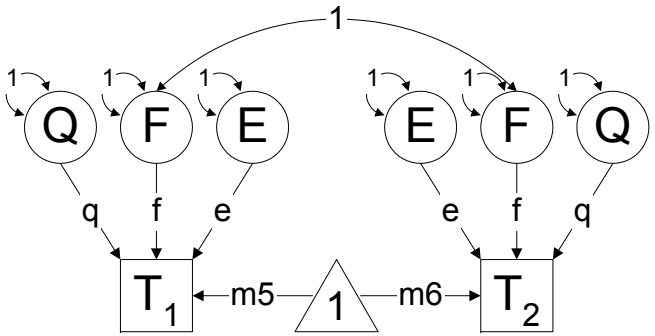
	DZibd2	DZibd1	DZibd0
piq	.60	.27	.15

	All correlations equal		
	χ^2	df	p
piq	13.07	2	.000

DZ by IBD status







- Variance = Q + F + E
- Covariance = πQ + F + E

```
G3: DZ IBD1 twins
 Data NInput=10
  Rectangular File=piqDZ.rec
  Labels fam id1 id2 piq1 piq2 ibd0mnr ibd1mnr ibd2mnr
    pihat picat
  Select if picat =1;
  Select piq1 piq2 ;
 Begin Matrices = Group 1;
  M Full nvar nvarx2 Free
  K Full 1 1                    ! correlation QTL effects
 End Matrices;
  Matrix M 110 110
  Matrix K .5
 Means M;
 Covariance
    F+Q+E | F+K@Q _
    F+K@Q | F+Q+E;
End
```

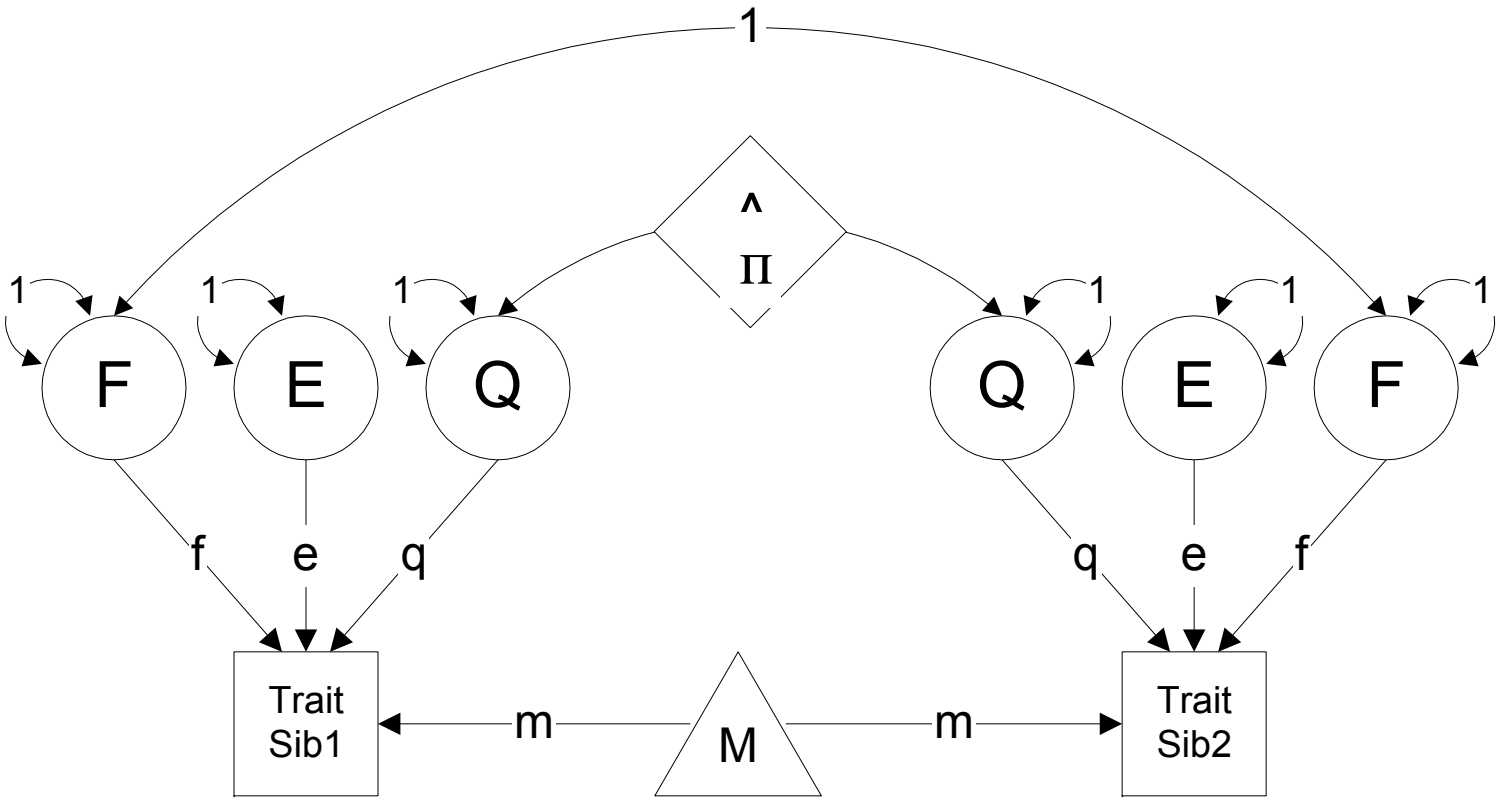
FEQmodel_DZibd.mx

Chi-square test for QTL + estimates

	Drop QTL		
	χ^2	df	p
piq	13.07	1	.000

	f^2	e^2	q^2
piq	.10 (.00-.27)	.43 (.32-.58)	.46 (.22-.67)

Linkage with Phi-Hat

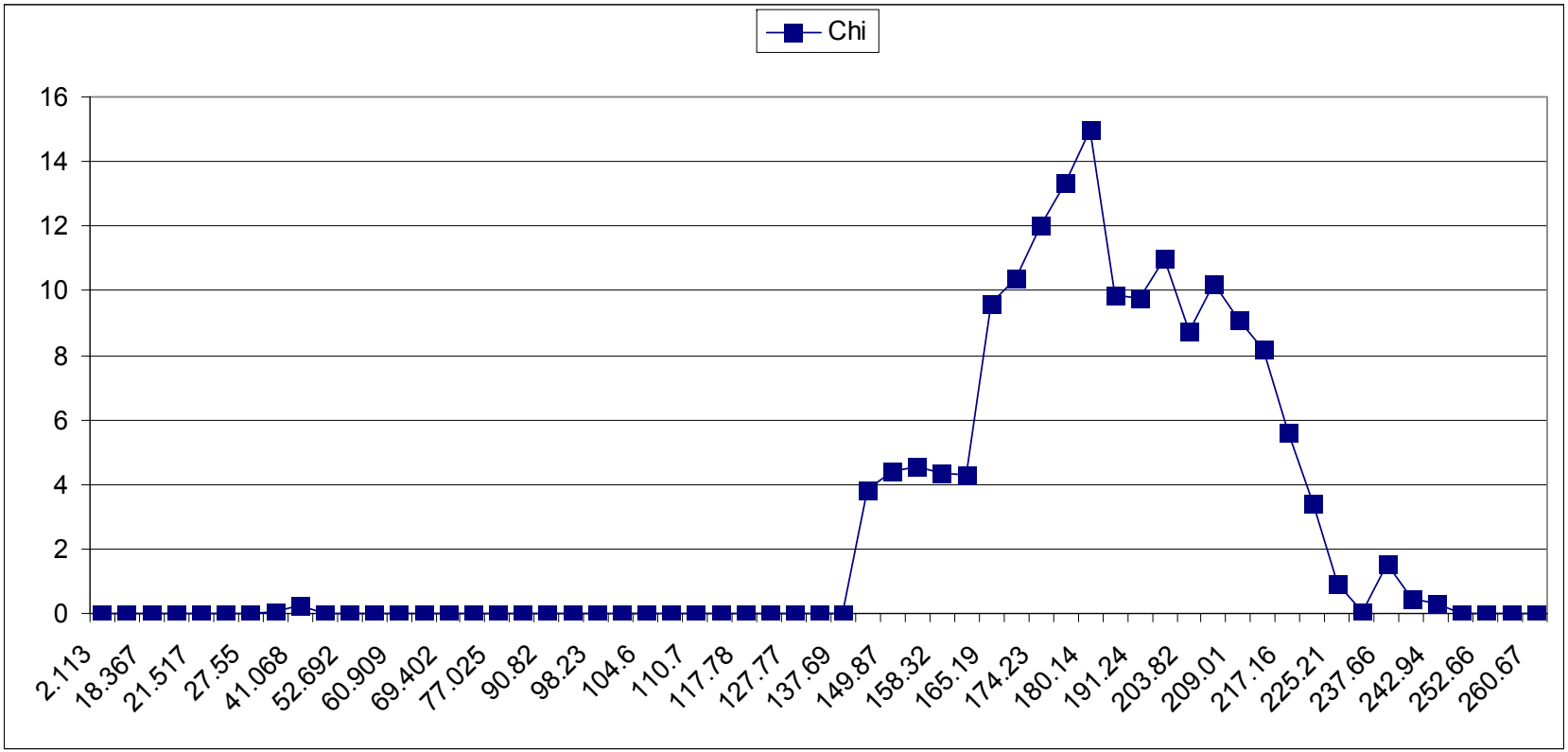


Definition Variables

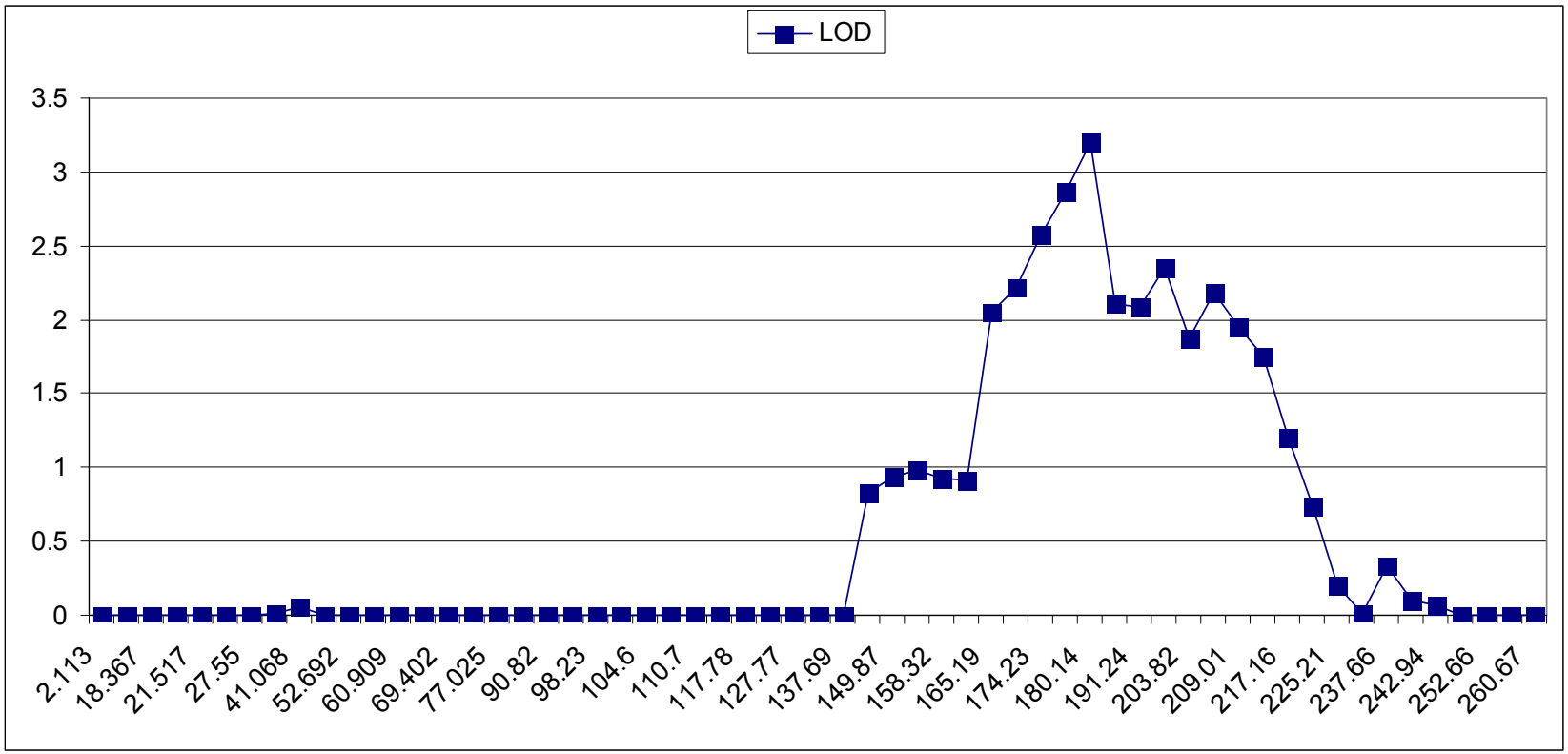
```
   Specify K    ibd0m1  ibd1m1  ibd2m1 ;
   Matrix H .5
   Matrix J 0 .5 1
   Start ..
  Begin Algebra;
   F= X*X';                    ! residual familial var
   E= Z*Z';                    ! unique environmental var
   Q= L*L';                    ! variance due to QTL
   V= F+Q+E;                   ! total variance
   T= F|Q|E;                   ! parameters in 1 matrix
   S= F%V| Q%V| E%V;           ! standardized var components
   P= J*K;                     ! estimate of pi-hat
  End Algebra;
Means  G| G ;
Covariance       F+Q+E | F+P@Q_
                 F+P@Q | F+Q+E ;
 Option Multiple Issat
End
```

FEQmodel_Pihat1_DZibd.mx

Pi-hat Results



LOD=(Univariate)$\Delta\chi^2/4.61$



Running a loop (Mx Manual page 52)

- Include a loop function in your Mx script
  - Analyze all markers consecutively
- At the top of the loop
  - `#loop $<number> start stop increment`
    - #loop $nr 1 59 1
- Within the loop
  - One file per chromosome, multiple markers
    - `Select piq1 piq2 ibd0m$nr ibd1m$nr ibd2m$nr`
  - One file per marker, multiple files
    - `Rectangular File =piq$nr.rec`
- At the end of the loop
  - `#end loop`

Outline

- Summary of yesterday afternoon
- Linkage in Merlin – Phi-Hat > Kate
- Linkage in Mx – Mixture > Meike

Mx vs *MERLIN*

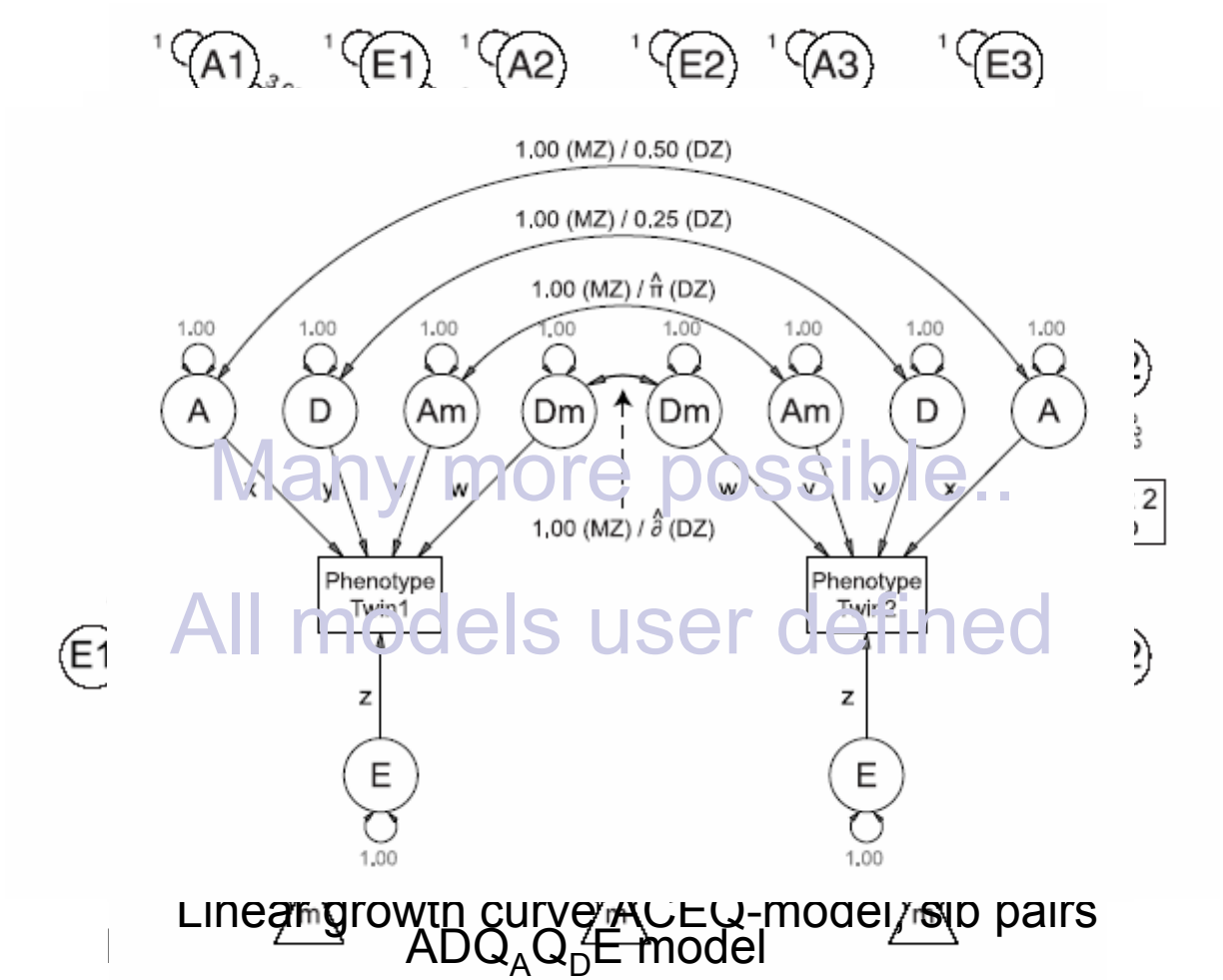
Mx

- Does not calculate IBDs
- Model specification nearly unlimited
  - multivariate phenotypes
  - Longitudinal modelling
  - Factor analysis
  - Sample heterogeneity testing
  - …
- No Graphical output

MERLIN

- Calculates IBDs
- Model specification relatively limited
- Some graphical output

Mx -



Linear growth curve ACEQ-model, sib pairs

ADQ_AQ_DE model

Trivariate model including covariates

Merlin Output (merlin.ibd)

- FAMILY ID1 ID2 MARKER P0 P1 P2
- 80020 3 3 2.113 0.0 0.0 1.0
- 80020 4 3 2.113 1.0 0.0 0.0
- 80020 4 4 2.113 0.0 0.0 1.0
- 80020 12 3 2.113 0.0 1.0 0.0
- 80020 12 4 2.113 0.0 1.0 0.0
- 80020 12 12 2.113 0.0 0.0 1.0
- 80020 11 3 2.113 0.0 1.0 0.0
- 80020 11 4 2.113 0.0 1.0 0.0
- 80020 11 12 2.113 0.32147 0.67853 0.00000
- 80020 11 11 2.113 0.0 0.0 1.0
- 80020 3 3 12.572 0.0 0.0 1.0
- 80020 4 3 12.572 1.0 0.0 0.0
- 80020 4 4 12.572 0.0 0.0 1.0
- 80020 12 3 12.572 0.0 1.0 0.0
- 80020 12 4 12.572 0.0 1.0 0.0
- 80020 12 12 12.572 0.0 0.0 1.0
- 80020 11 3 12.572 0.0 1.0 0.0
- 80020 11 4 12.572 0.0 1.0 0.0
- 80020 11 12 12.572 0.70372 0.29628 0.00000

Merlin Output (merlin.ibd)

- Merlin will output IBD estimates for all possible pairs that can be created within a single family.
- Some of these IBD estimates are invariant

for example:

spouses will always be IBD = 0

parent-offspring relations will always be IBD = 1

Merlin Output (merlin.ibd)

- In some cases, IBD estimates are not invariant by default, but may still follow an a priori pattern (i.e. for sibling pairs the probabilities for sharing 0, 1, or 2 alleles IBD will be ¼, ½, and ¼ respectively)

>> The latter will happen when one or both members are not genotyped, or are genotyped for only a very small portion of all available genotypes.

Mx Input (piqibd.rec)

- 80020 11 12 118 112 0.32147 0.67853 0 0.70372 0.29628 0 1 0 0 0.99529 0.00471 0 1 0 0 0.27173 0.72827 0 0.25302 0.74171 0.00527 0.03872 0.96128 0 0.02434 0.97566 0 0.01837 0.98163 0 0.01077 0.96534 0.02389 0.01976 0.98024 0 0.02478 0.97522 0 0.01289 0.98711 0 0.01124 0.98876 0 0.00961 0.92654 0.06385 0.01855 0.98145 0 0.04182 0.95818 0 0.03635 0.96365 0 0.03184 0.85299 0.11517 0.00573 0.22454 0.76973 0.00229 0.13408 0.86363 0.00093 0.07687 0.9222 0 0.00209 0.9979 0 0.00221 0.99779 0.00002 0.00829 0.99169 0.00065 0.09561 0.90374 0.01589 0.98411 0 0.00991 0.99009 0 0.00443 0.99557 0 0.01314 0.98686 0 0.44616 0.55384 0 0.68628 0.31372 0 1 0 0 0.98957 0.01043 0 0.98792 0.01208 0 0.97521 0.02479 0 1 0 0 1 0 0 0.43647 0.55668 0.00685 0.28318 0.71682 0 0.14261 0.83132 0.02607 0.13582 0.86418 0 0.1056 0.8944 0 0.03629 0.96371 0 0.00279 0.27949 0.71772 0.00143 0.12575 0.87282 0.00011 0.02912 0.97078 0.00001 0.00592 0.99407 0.00002 0.00703 0.99295 0.00012 0.02351 0.97637 0.00064 0.06857 0.93078 0.00139 0.24954 0.74907 0.00784 0.99216 0 0.01713 0.94333 0.03954 0.057 0.943 0 0.05842 0.91425 0.02733 0.03722 0.96278 0 0.03722 0.96278 0
- 80030 12 11 121 127 0.05559 0.94441 0 0.07314 0.80951 0.11736 0.15147 0.84853 0 0.18374 0.81626 0 0.29586 0.70414 0 1 0 0 0.99416 0.00584 0 0.97643 0.02343 0.00014 1 0 0 1 0 0 0.9949 0.0051 0 1 0 0 0.94805 0.05195 0 1 0 0 0.95133 0.04864 0.00003 0.5887 0.4113 0 0.1536 0.8464 0 0.00204 0.10279 0.89517 0.00008 0.0541 0.94582 0.00026 0.07795 0.92179 0.00438 0.43379 0.56184 0.01809 0.98191 0 0.02748 0.97252 0 0.01871 0.98129 0 0.01907 0.98093 0 0.02263 0.97737 0 0.00829 0.442 0.54971 0.00066 0.13393 0.86541 0.00216 0.13426 0.86358 0.00138 0.08847 0.91015 0.0027 0.12535 0.87195 0.0035 0.21603 0.78047 0.02032 0.49739 0.48228 0.05 0.95 0 0.06282 0.92949 0.00769 0.06502 0.92616 0.00882 0.0801 0.9199 0 0.08891 0.91109 0 0.08646 0.91354 0 0.0813 0.9187 0 0.08568 0.91432 0 0.2608 0.7392 0 0.29967 0.70033 0 0.36423 0.63577 0 0.45359 0.53993 0.00649 0.48542 0.51458 0 1 0 0 1 0 0 0.48916 0.50519 0.00566 0.38395 0.61605 0 0.08177 0.91823 0 0.06985 0.90434 0.02581 0.01758 0.98242 0 0.00242 0.99758 0 0.00914 0.99086 0 0.04127 0.95873 0 0.05606 0.93267 0.01127 0.06201 0.93799 0 0.06201 0.93799 0

fam id1 id2 piq1 piq2 ibd0m1 ibd1m1 ibd2m1 ibd0m2 ibd1m2 ibd2m2 ….

phenotypes *ibd probabilities to calculate pihats at different locations*

Once you have your data ….

- Incorporate QTL effects in ACE/ADE models
- ‘Simple’ extension of path models and Mx scripts

Alternative way to model linkage

Rather than categorize or calculating pi-hat, we can fit three models (for ibd=0, 1, or 2) to the data and weight each model by its corresponding IBD probability for a pair of siblings:

Full information approach aka

Weighted likelihood or

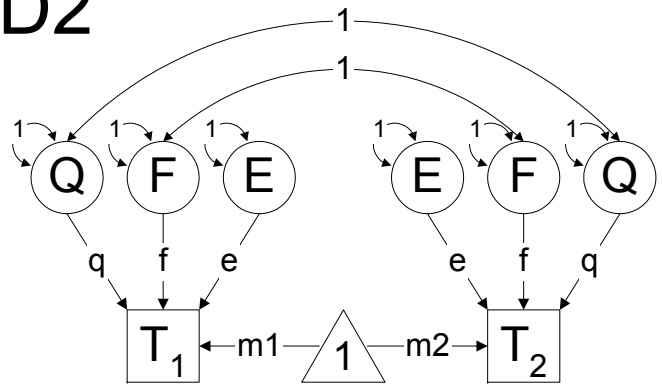
Mixture distribution approach

Mixture Distribution Approach

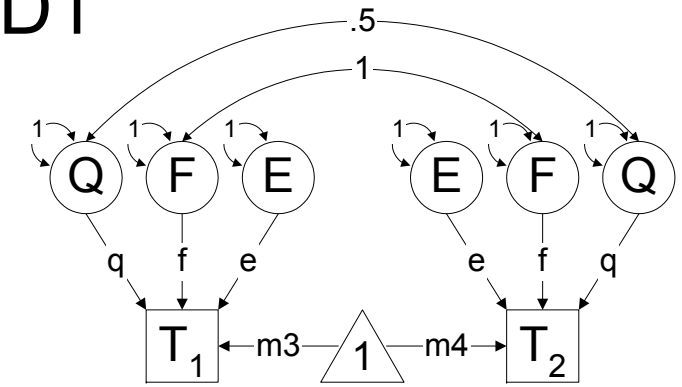
In the mixture distribution approach to linkage, we fit three models (for IBD=0, IBD=1, IBD=2) for each sib pair, each weighted by their relative IBD probabilities.

DZ by IBD status -> QFE

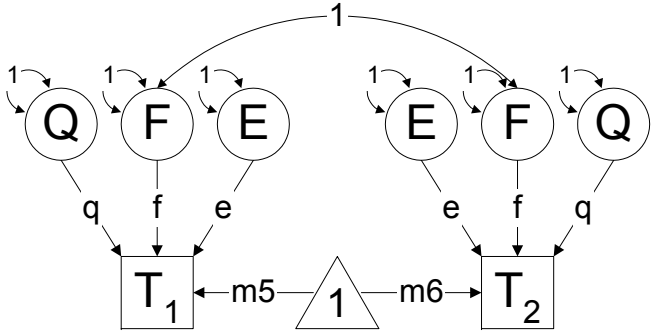
IBD2



IBD1



IBD0



- Variance = Q + F + E
- Covariance = πQ + F + E

```
#define nvar 1
#NGroups 1

DZ / SIBS genotyped
 Data NInput=182 Maxrec=1500 NModel=3
 Rectangular File=piqibd.txt
  Labels fam id1 id2 piq1 piq2
 ibd0m1 ibd1m1 ibd2m1 ibd0m2 ibd1m2 ibd2m2 ....
 ibd0m59 ibd1m59 ibd2m59
  Select  piq1 piq2  ibd0m1  ibd1m1  ibd2m1 ;
  Definition    ibd0m1  ibd1m1  ibd2m1 ;
 Begin Matrices;
  X Lower nvar nvar free ! residual familial F
  Z Lower nvar nvar free ! unshared environment E
  L Full nvar 1 free     ! qtl effect Q
  G Full 1 nvar free     ! grand means
  H Full 1 1             ! scalar, .5
  K Full 3 1             ! IBD probabilities (Merlin)
  U Unit 3 1             ! to extend means
 End Matrices;
```

```
  Specify K    ibd0m1  ibd1m1  ibd2m1 ;
  Matrix H .5
  Start ..
 Begin Algebra;
  F= X*X';                    ! residual familial var
  E= Z*Z';                    ! unique environmental var
  Q= L*L';                    ! variance due to QTL
  V= F+Q+E;                   ! total variance
  T= F|Q|E;                   ! parameters in 1 matrix
  S= F%V| Q%V| E%V;           ! standardized var components
 End Algebra;
 Means   U@G| U@G ;
 Covariance
  F+E+Q | F       _
  F     | F+E+Q   _           ! IBD 0 Covariance matrix
  F+E+Q | F+H@Q   _
  F+H@Q | F+E+Q   _           ! IBD 1 Covariance matrix
  F+E+Q | F+Q     _
  F+Q   | F+E+Q;              ! IBD 2 Covariance matrix
 Weights K;                   ! IBD probabilities
 Option Multiple Issat
End
```

Practical Mixture

- Mx script: mixture_piq_Prac.mx

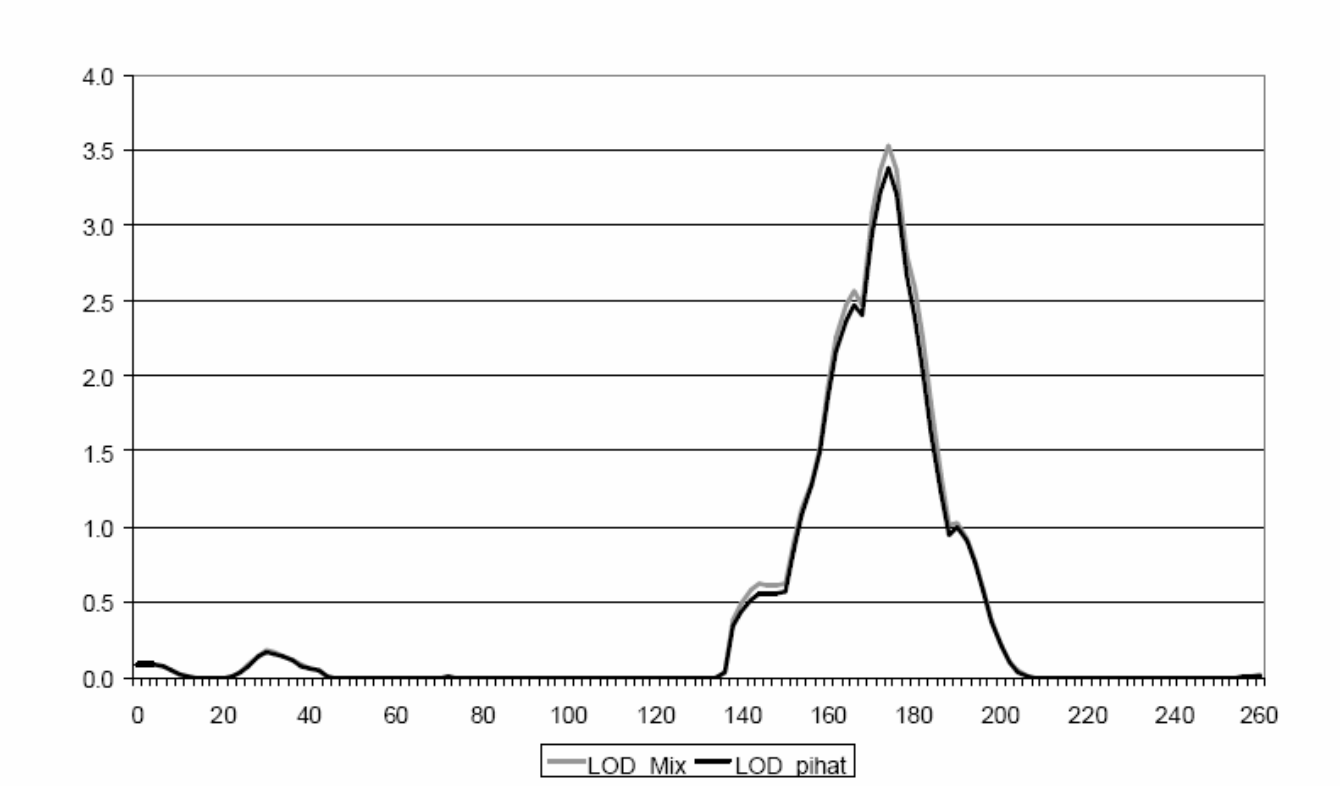
- Fill in ????
- Choose a position, run model
- Calculate lod-score

faculty:\meike\2007\mixture\mixture_piq_Prac.mx

Pi-hat versus Mixture

- Pi-hat simple with large sibships
  - Solar, Genehunter, etc.
- Pi-hat shows substantial bias with missing data
  - Example: Pi-hat=.4 may result from
    - ibd0=.33 ibd1=.33 ibd2=.33
    - ibd1=0 ibd2=.5
  - Thus mixture retains all information > more power
  - Pi-hat does not

Results Phi-Hat vs. Mixture



http://www.psy.vu.nl/mxbib/

Individual Likelihoods

- Mx allows to output the contribution to the -2ll per family:
- Raw data
- Options Mx%p= file.out
- Will output for each case in the data the contribution to the -2ll as well as z-score statistic and Mahalanobis distance

1	2	3	4	5	6	7	8	9
9.000000000000000	7.336151039930395	1.540683866365682	8.343722785165869E-02	1	2	0	000	1
10.00000000000000	9.851302691037933	4.055835517473221	1.130602365719245	2	2	0	000	1
12.00000000000000	7.143777518583584	1.348310345018871	-3.614906755842623E-02	3	2	0	000	1

- 1. first definition variable (wise to use a case identifier)
- 2. -2lnL: likelihood function for vector of observations
- 3. square root of the Mahalanobis distance
- 4. estimated z-score
- 5. number of the observation in the active (i.e. post selection) dataset
- 6. number of data points in the vector (i.e. the family size if it is a pedigree with one variable per family member
- 7. number of times the log-likelihood was found to be incalculable during optimization
- 8. 000 if likelihood was able to be evaluated at the solution, or 999 if it was incalculable
- 9. model number if there are multiple models requested with the NModel argument

Practical %p

- Adjust pihat_piq1.mx to run at position with highest lod score
- Select variable fam, #define fam as first definition variable
- Run QFE model with: Options Mx%p=QFE.dat
- Run FE model with: Options Mx%p=FE.dat
- Import the two dat files in excell (contribution to LL.xls) select the first two columns of each dat file.
- Subtract the -2ll per family
- Sort the file on the difference in -2ll
- Produce a graph

share:\h.maes\MxLinkage\

%p Viewer

- Java applet from QIMR to view the %p output in a convenient way
- Open viewdist.jar,open QFE.dat

Using MZ twins in linkage

- An MZ pair will not contribute to your linkage signal
  - BUT correctly including MZ twins in your model allows you to partition F in A and C or in A and D
  - AND if the MZ pair has a (non-MZ) sibling the ‘MZ-trio’ contributes more information than a regular (DZ) sibling pair – but less than a ‘DZ-trio’
  - MZ pairs that are incorrectly modelled lead to spurious results

From Merlin to Mx

- Different ways to go about this
  - Shell or Perl scripts in Unix/Linux 👍☹
  - SPSS, SAS, R etc 👎☺
  - alsort (for pairwise data) 👍☺
    - Takes an all-possible pairs approach rather than a full sib-ship approach
      - If a family has a sibship of 2 then 1 pair
      - If a family has a sibship of 3 then 3 pairs
      - If a family has a sibship of 4 then 6 pairs
    - You can run alsort and then convert to a full sib-ship approach

alsort.exe

```
Usage:
alsort <inputfile> <outpfile> [-vfpm] [-c] [-i] [-t]
   [-x <id1> ...]

-v    Verbose (implies -vfpm)
-vf   Print family ID list
-vp   Print marker positions
-vm   Print missing p-values
-c    Create output file per chromosome
-i    Include 'self' values (id1=id2)
-x    Exclude list; id-values separated by spaces
-t    Write tab as separator character
```

Practical alsort.exe

- Open a dos prompt
- Go to directory where alsort.exe is
- And type
  - alsort merlin.ibd sort.txt –c –x 3 4 –t
  - (3 & 4 are id's for parents)

share:\h.maes\MxLinkage\