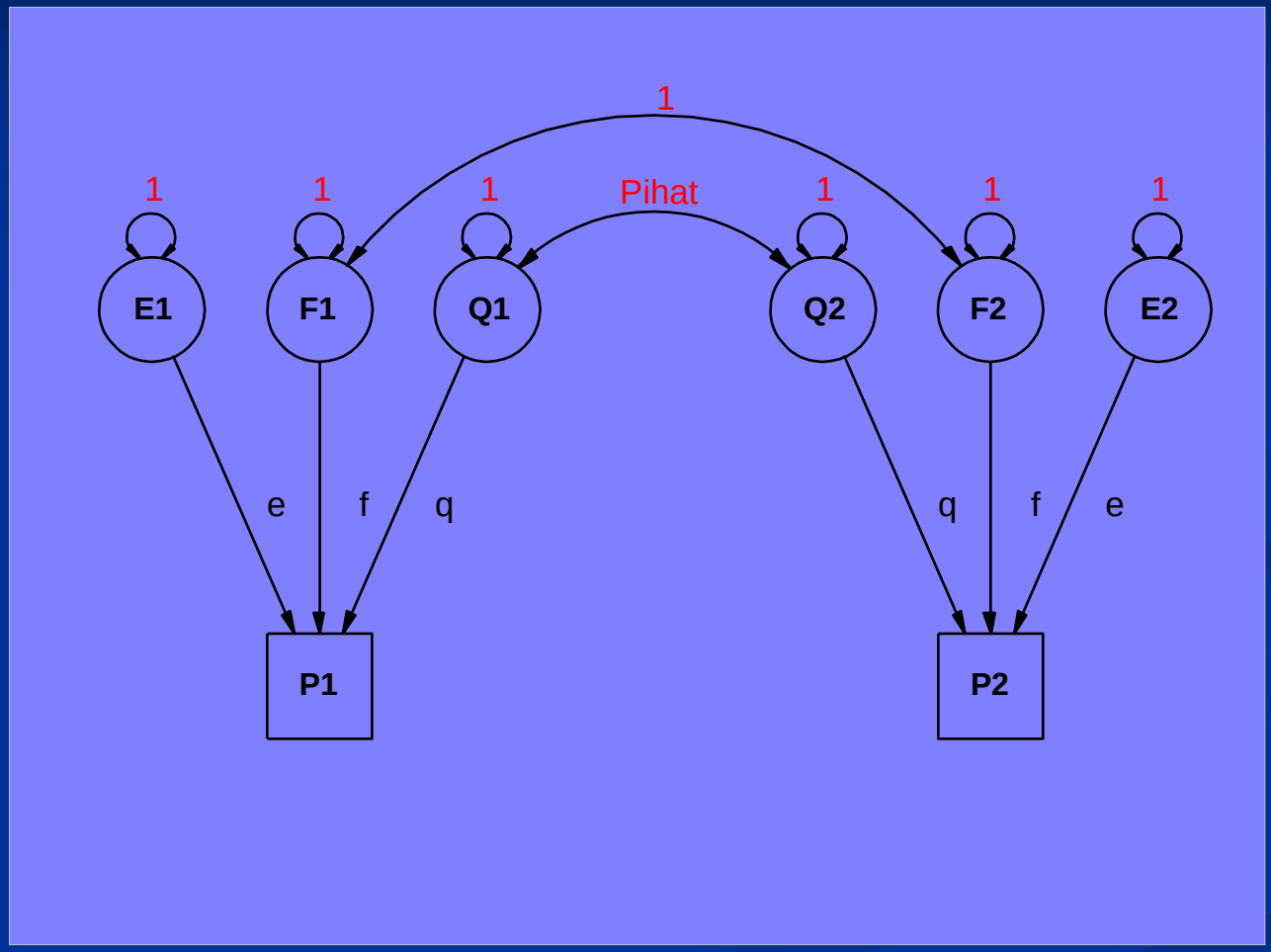
Linkage in Selected Samples

Boulder Methodology Workshop
2005

Michael C. Neale
Virginia Institute for Psychiatric & Behavioral Genetics

Basic Genetic Model

Pihat = p(IBD=2) + .5 p(IBD=1)

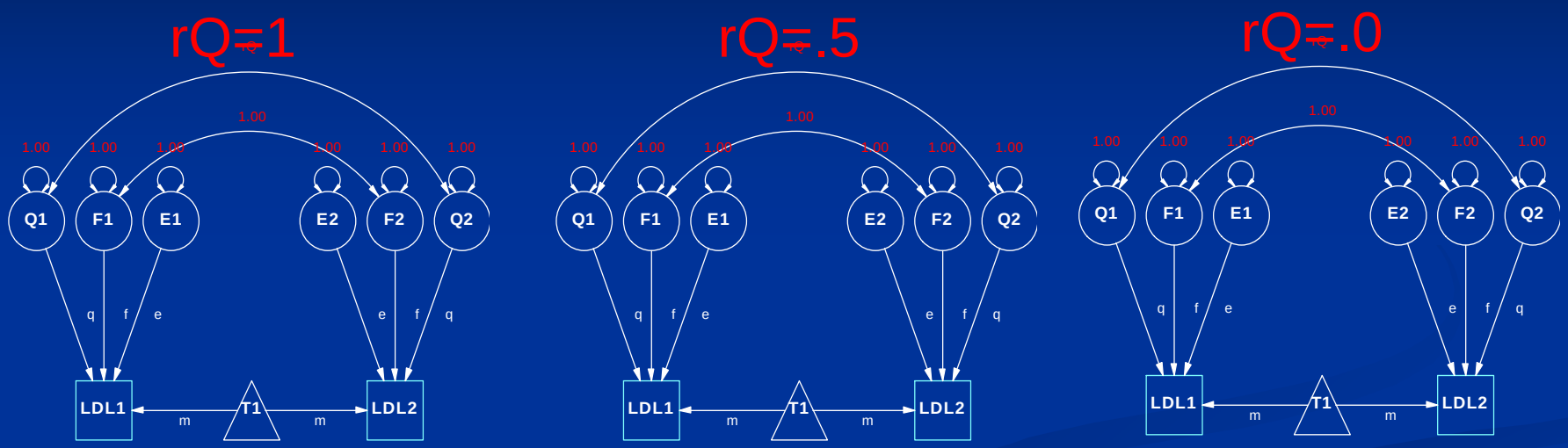


Q: QTL Additive Genetic F: Family Environment E: Random Environment

3 estimated parameters: q, f and e Every sibship may have different model

Mixture distribution model

Each sib pair i has different set of WEIGHTS



$weight_j$ x Likelihood under model j

p(IBD=2) x P(LDL1 & LDL2 | rQ = 1)

p(IBD=1) x P(LDL1 & LDL2 | rQ = .5)

p(IBD=0) x P(LDL1 & LDL2 | rQ = 0)

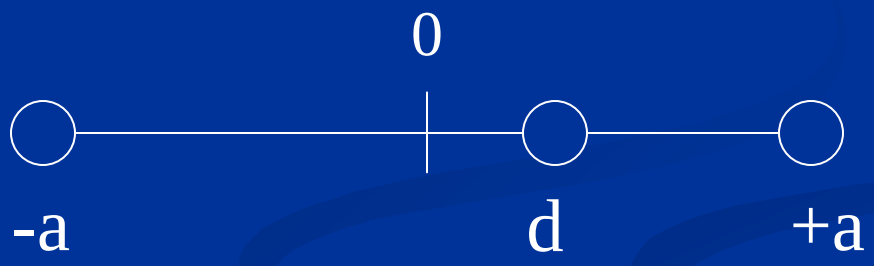
Total likelihood is sum of weighted likelihoods

QTL's are factors

- Multiple QTL models possible, at different places on genome
- A big QTL will introduce non-normality
  - Introduce mixture of means as well as covariances (27ish component mixture)
- Mixture distribution gets nasty for large sibships

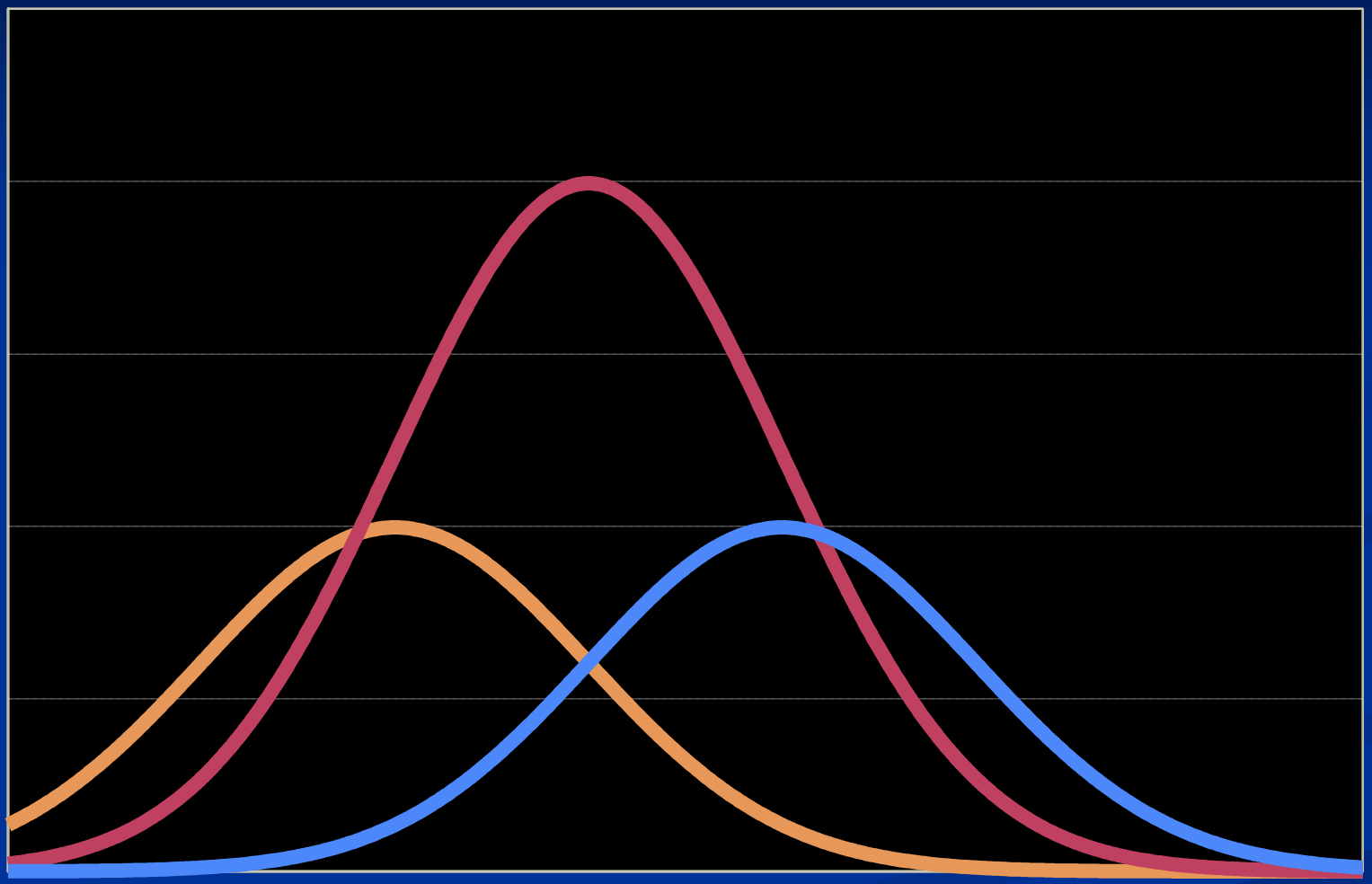
Biometrical Genetic Model

	Genotype means
AA	m + a
Aa	m + d
aa	m – a



Courtesy Pak Sham, Boulder 2003

Mixture of Normal Distributions

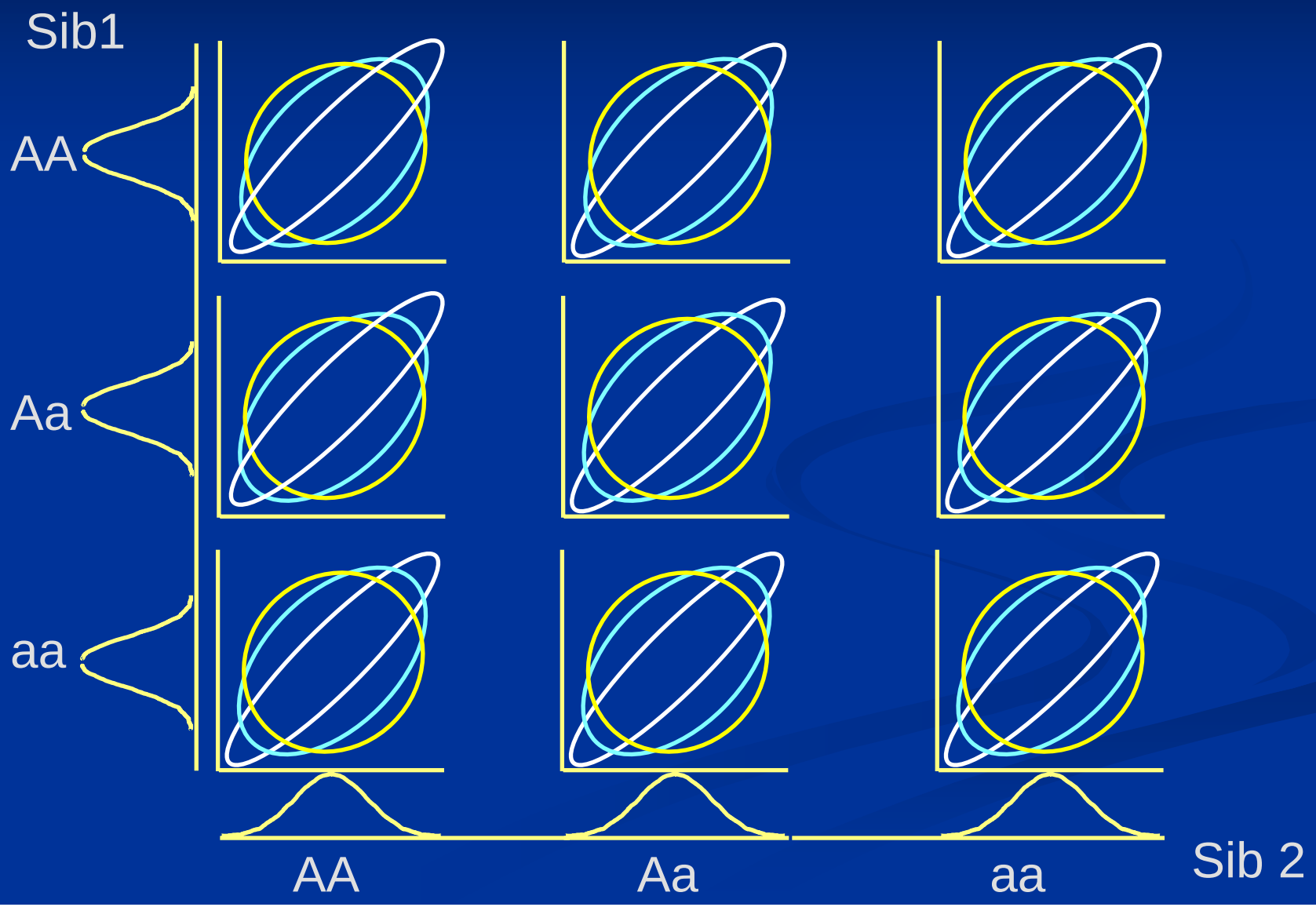


Equal variances, Different means and different proportions according to allele frequencies

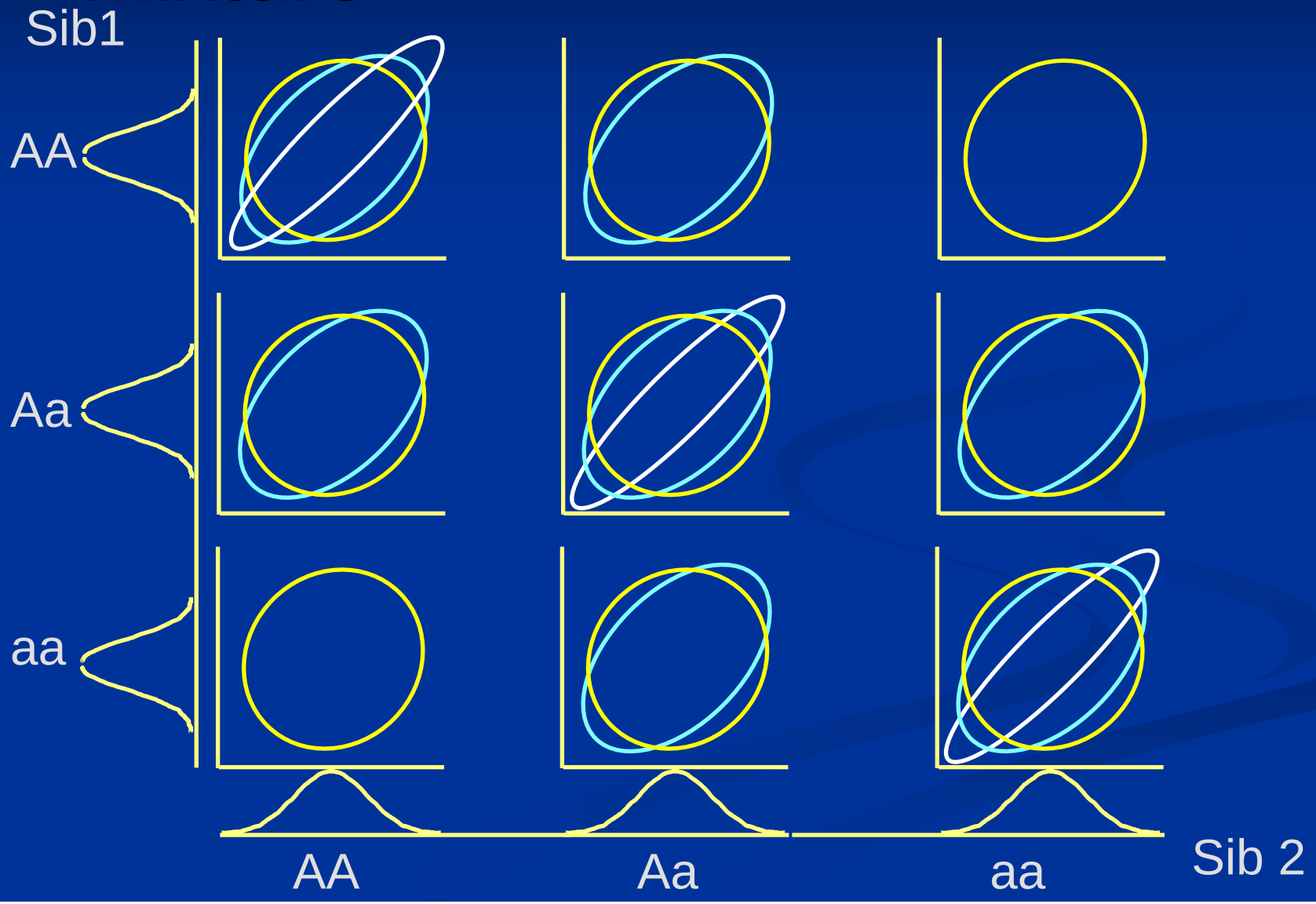
Implementing the Model

- Estimate QTL allele frequency p
- Estimate distance between homozygotes 2a
- Compute QTL additive genetic variance as
  - $2pq[a+d(q-p)]^2$
- Compute likelihood conditional on
  - IBD status
  - QTL allele configuration of sib pair (IBS)

27 Component Mixture



19 Possible Component Mixture



Results of QTL Simulation

3 Component vs 19 Component

Parameter	True	3 Component	19 Component
Q	0.4	0.414	0.395
A	0.08	0.02	0.02
E	0.6	0.56	0.58
Test Q=0 (Chisq)	---	13.98	15.88

200 simulations of 600 sib pairs each GASP http://www.nhgri.nih.gov/DIR/IDRB/GASP/

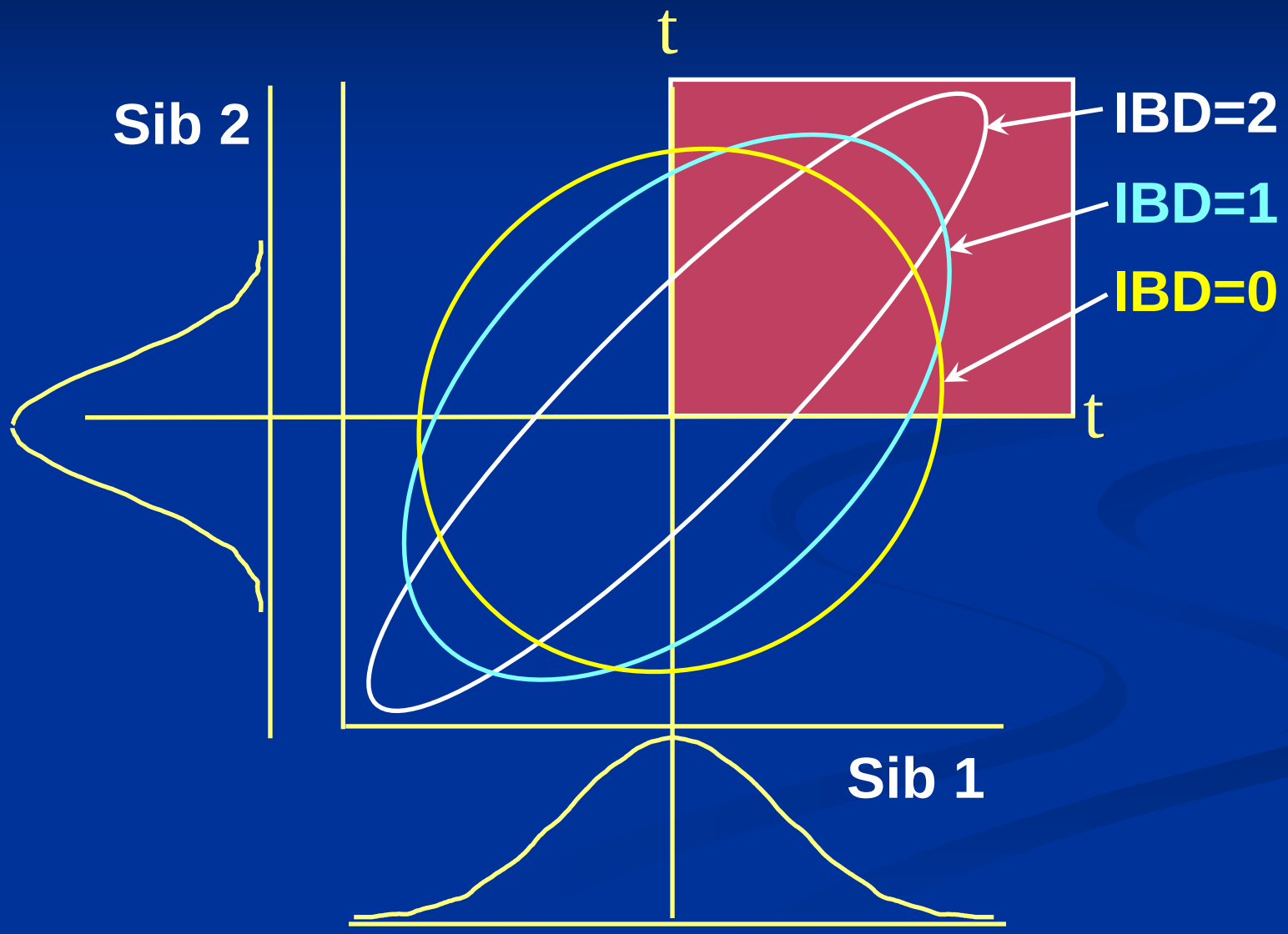
Information in selected samples

Concordant or discordant sib pairs

- Deviation of pihat from .5
  - Concordant high pairs > .5
  - Concordant low pairs > .5
  - Discordant pairs < .5
- How come?

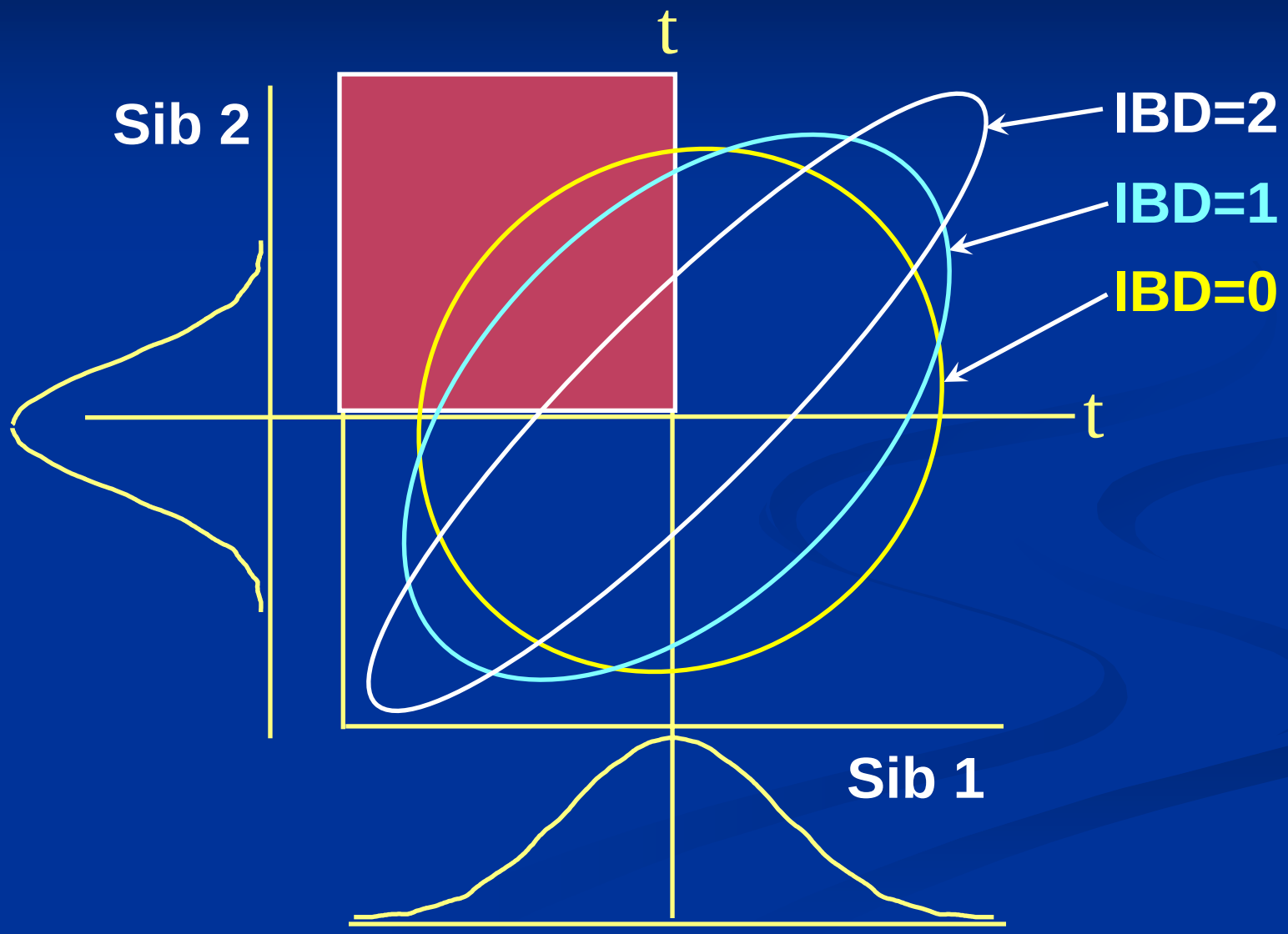
Pihat deviates > .5 in ASP

Larger proportion of IBD=2 pairs

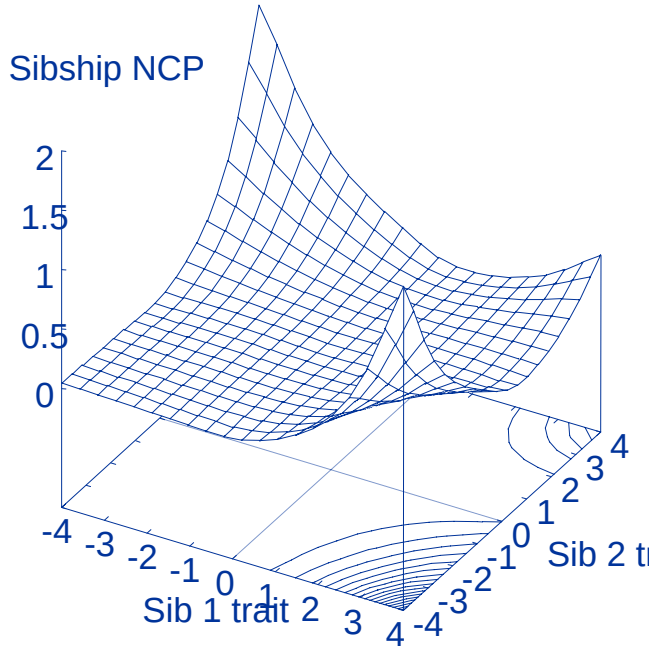


Pihat deviates <.5 in DSP's

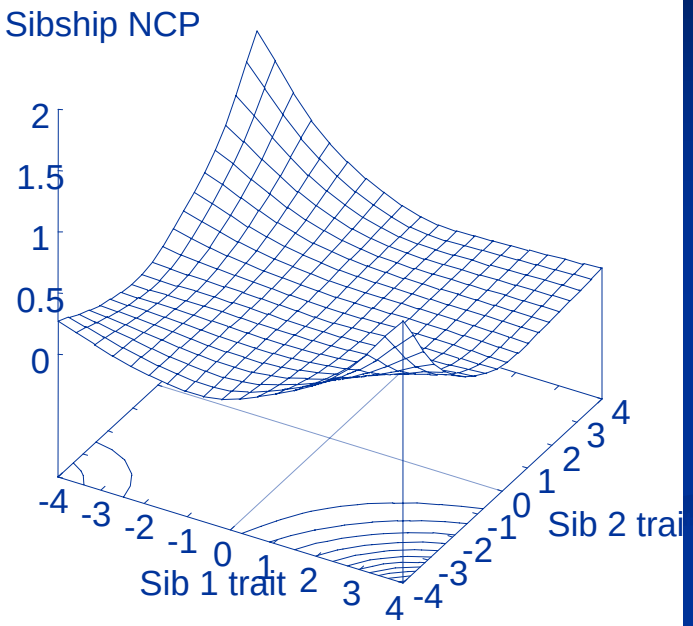
Larger proportion of IBD=0 pairs



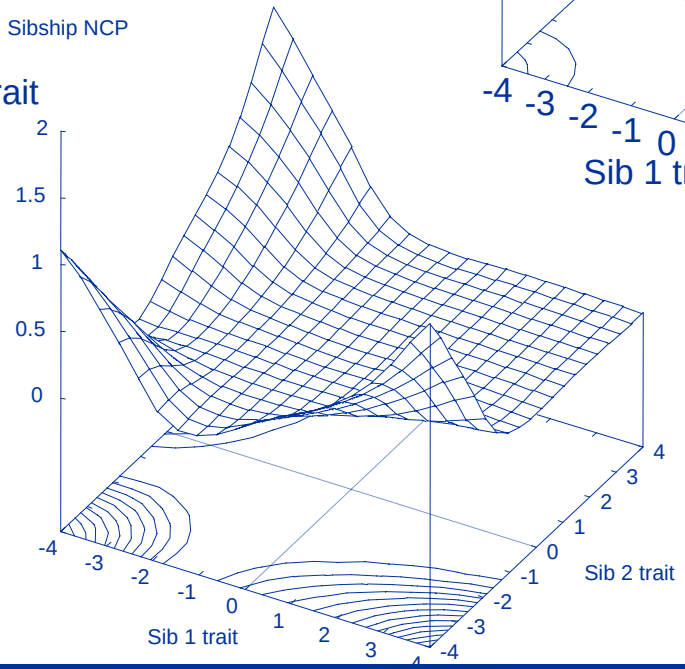
Sibship informativeness : sib pairs



dominance



unequal allele frequencies



rare recessive

Courtesy Shaun Purcell, Boulder Workshop 03/03

Two sources of information

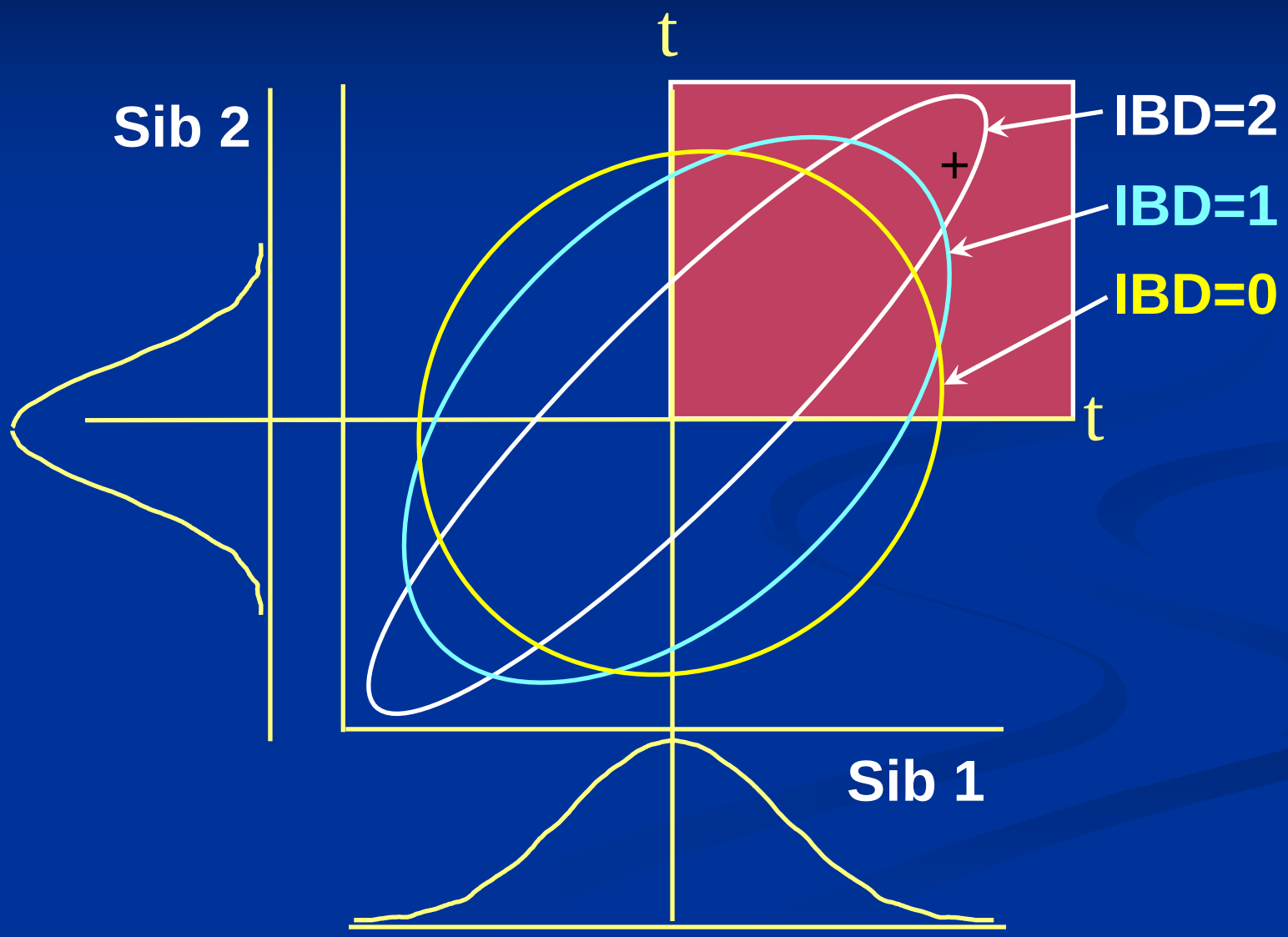
Forrest & Feingold 2000

- Phenotypic similarity
  - IBD 2 > IBD 1 > IBD 0
  - Even present in selected samples
- Deviation of pihat from .5
  - Concordant high pairs > .5
  - Concordant low pairs > .5
  - Discordant pairs < .5
- These sources are independent

Implementing F&F

- Simplest form test mean pihat = .5
- Predict amount of pihat deviation
  - Expected pihat for region of sib pair scores
  - Expected pihat for observed scores
- Use multiple groups in Mx

Predicting Expected Pihat deviation



Expected Pihats: Theory

- IBD probability conditional on phenotypic scores x_1, x_2

- $E(\text{pihat}) = p(IBD=2|(x_1,x_2)) + .5p(IBD=1|(x_1,x_2))$

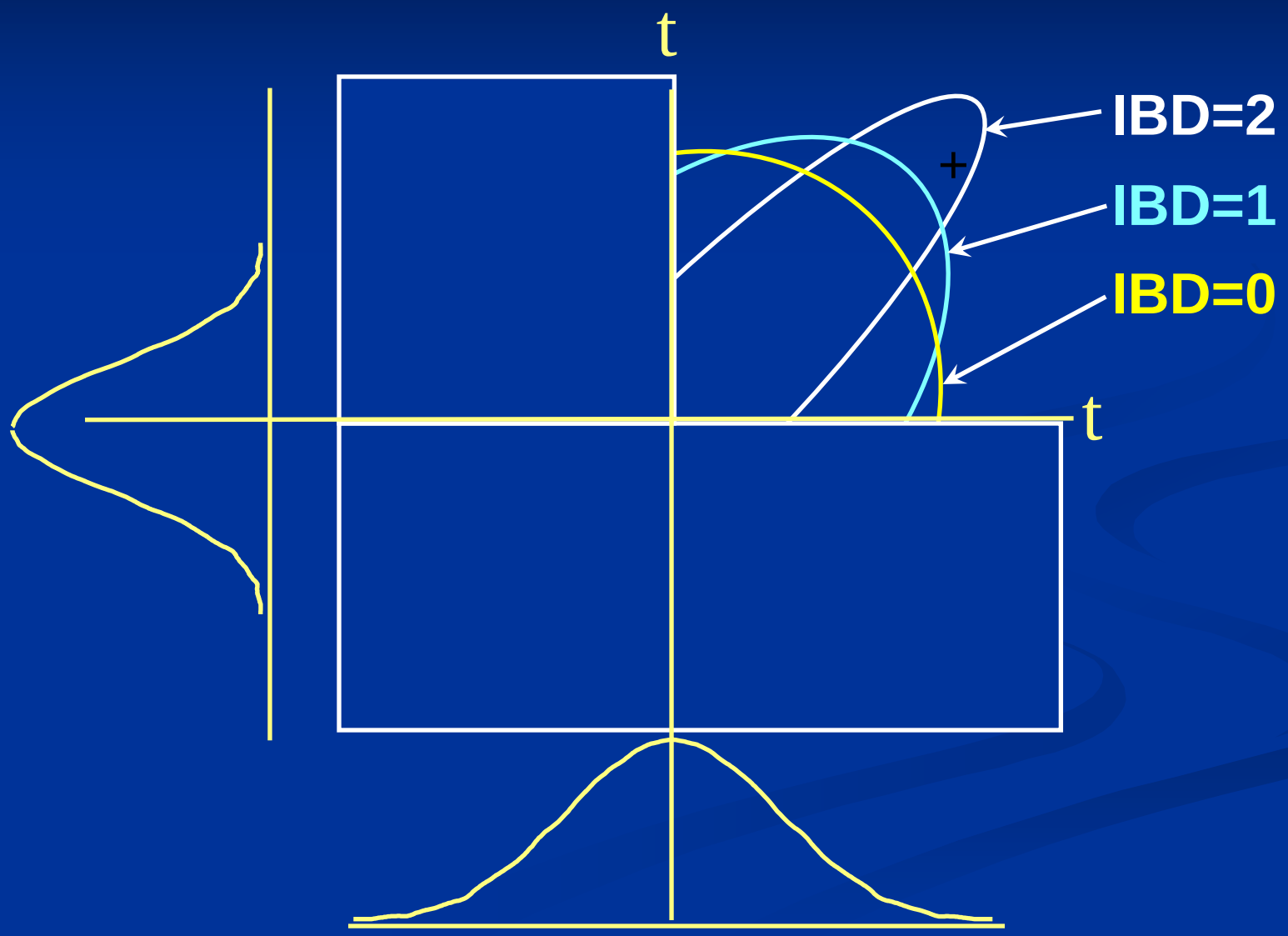
$$\frac{}{p(IBD=2|(x_1,x_2)) + p(IBD=1|(x_1,x_2)) + p(IBD=0|(x_1,x_2))}$$

- $p(IBD=2|(x_1,x_2)) = \text{☠}_{IBD=2}(x_1,x_2)$ /
- $[\text{☠}_{IBD=2}(x_1,x_2) + 2\text{☠}_{IBD=1}(x_1,x_2) + \text{☠}_{IBD=0}(x_1,x_2)]$

Expected Pihats

- Compute Expected Pihats with pdfnor
- \pdfnor(X_M_C)
  - Observed scores X (row vector 1 x nvar)
  - Means M (row vector)
  - Covariance matrix C (nvar x nvar)

How to measure covariance?



Ascertainment

- Critical to many QTL analyses
- Deliberate
  - Study design
- Accidental
  - Volunteer bias
  - Subjects dying

Exploiting likelihood

- Correction not always necessary
  - ML MCAR/MAR

- Simulate bivariate normal data X,Y
- Sigma = 1 .5
- .5 1
- Mu = 0, 0
-
- Make some variables missing
- Generate independent random normal variable, Z, if Z>0 then Y miss
- If X>0 then Y missing
- If Y>0 then Y missing
-
- Estimate elements of Sigma & Mu
- Constrain elements to population values 1,.5, 0 etc
- Compare fit
- Ideally, repeat multiple times and see if expected 'null' distribution emerge

Results of simulation

Population covariance 1 .5 1 Means 0, 0

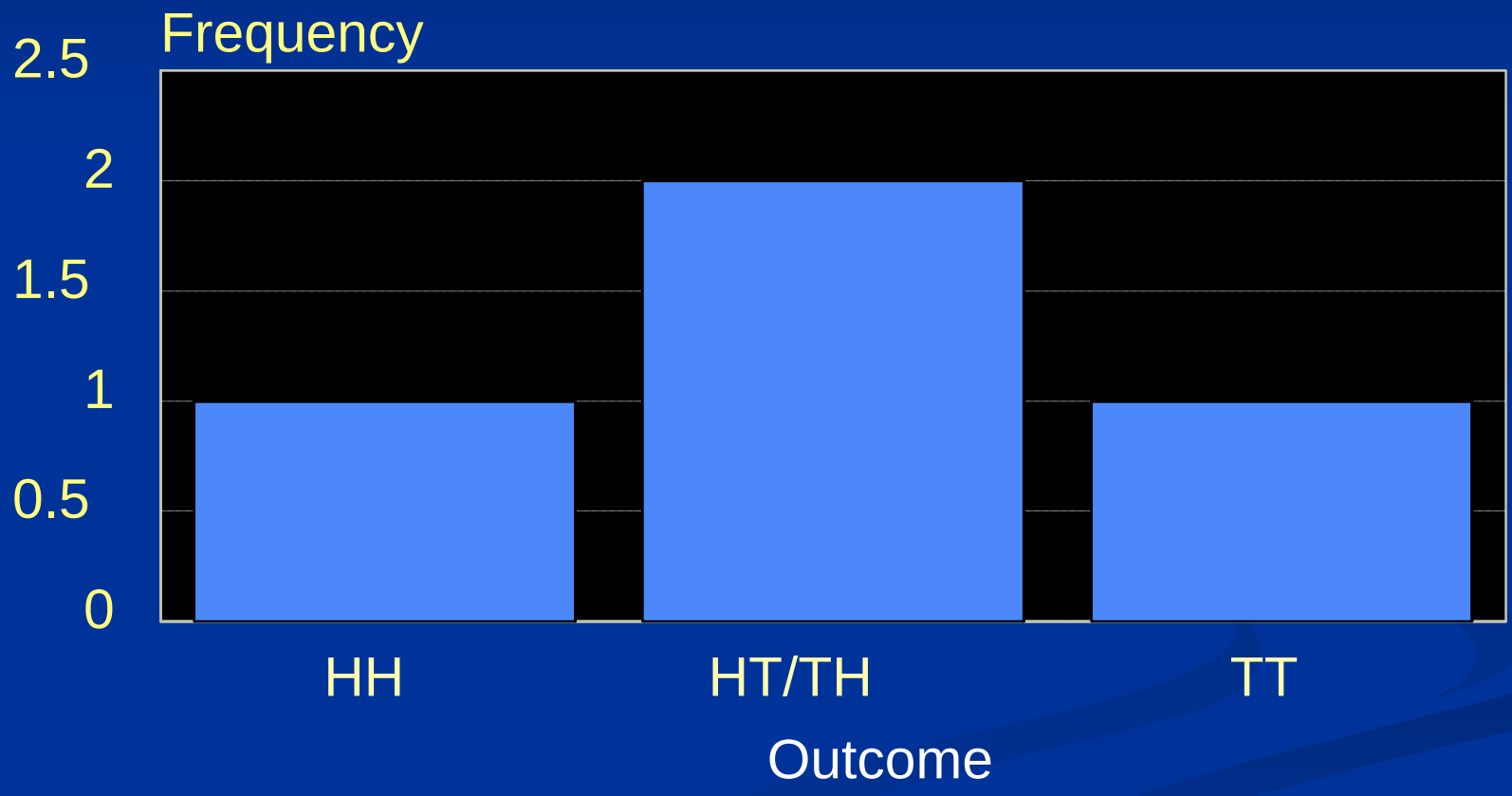
Missingness	mean x	mean y	var x	cov xy	var y	LR Chisq
MCAR (rand) MLE	-0.0116	-0.1	1.0505	0.4998	0.8769	6.492
sample	-0.0116	-0.0919	1.0505		0.8839	
MAR (on x) MLE	0.0048	0.0998	1.0084	0.4481	1.1025	5.768
sample	0.0014	0.4437	1.0084		0.9762	
NMAR (on y) MLE	-0.0204	0.6805	0.9996	0.1356	0.2894	227.262
sample	0.0448	0.7373	0.9996		0.2851	

Weighted likelihood approach

- Usual nice properties of ML remain
- Flexible
- Simple principle
  - Consideration of possible outcomes
  - Re-normalization
- May be difficult to compute

Example: Two Coin Toss

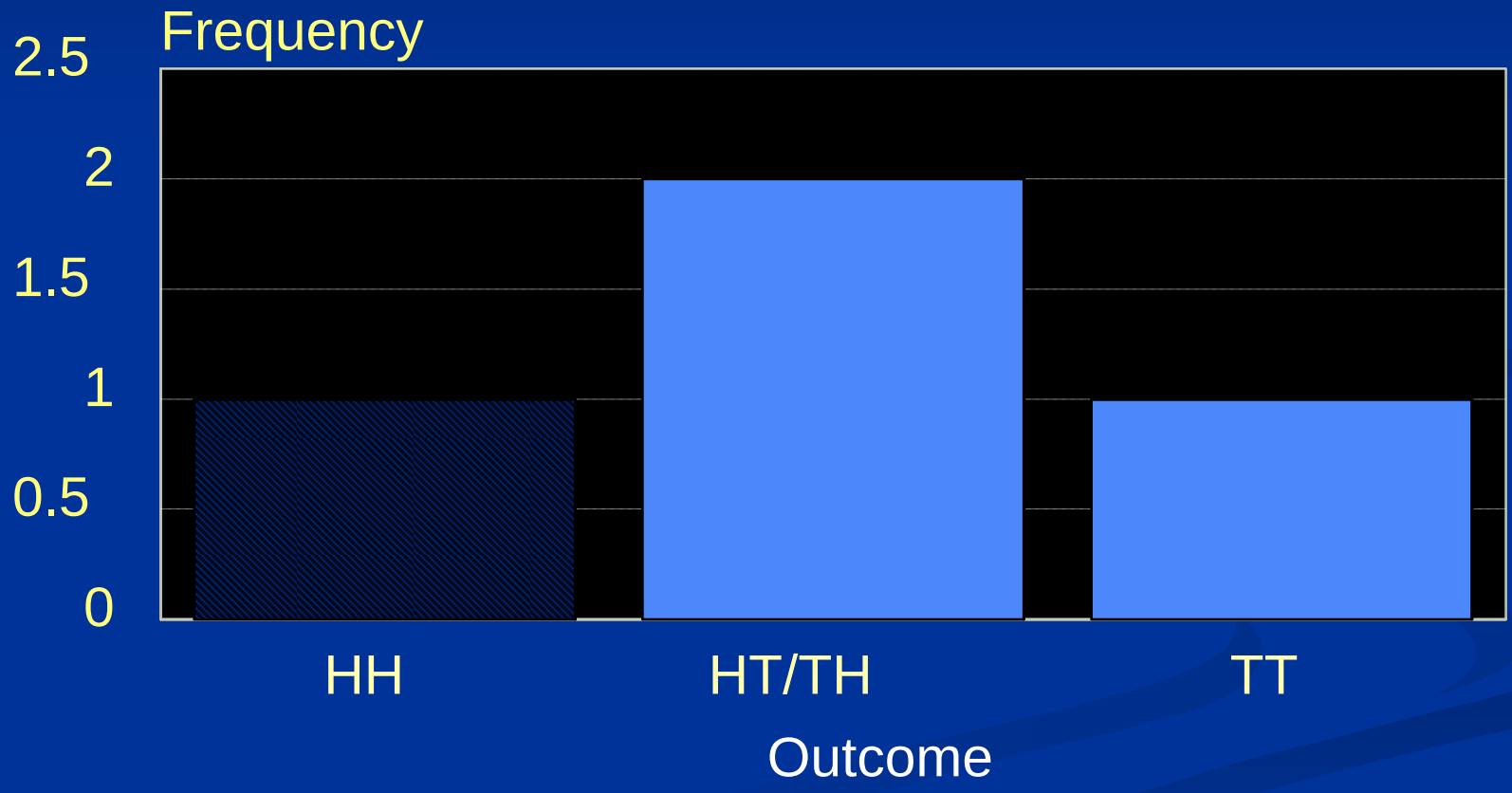
3 outcomes



Probability i = freq i / sum (freqs)

Example: Two Coin Toss

3 outcomes



Probability i = freq i / sum (freqs)

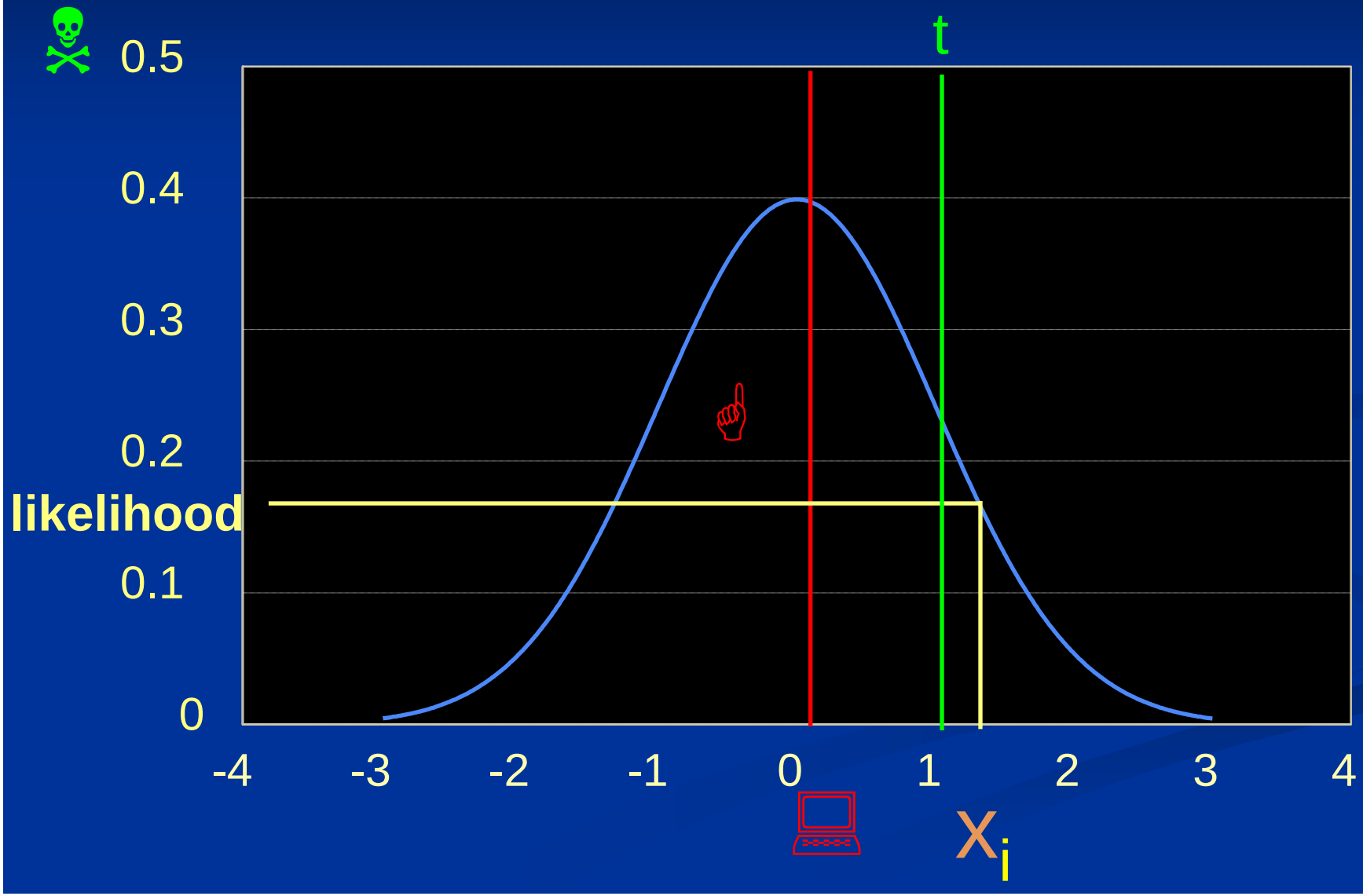
Non-random ascertainment

Example

- Probability of observing TT globally
  - 1 outcome from 4 = 1/4
- Probability of observing TT if HH is not ascertained
  - 1 outcome from 3 = 1/3
  - or 1/4 divided by '*Ascertainment Correction*' of 3/4 = 1/3

Correcting for ascertainment

Univariate case; only subjects > t ascertained

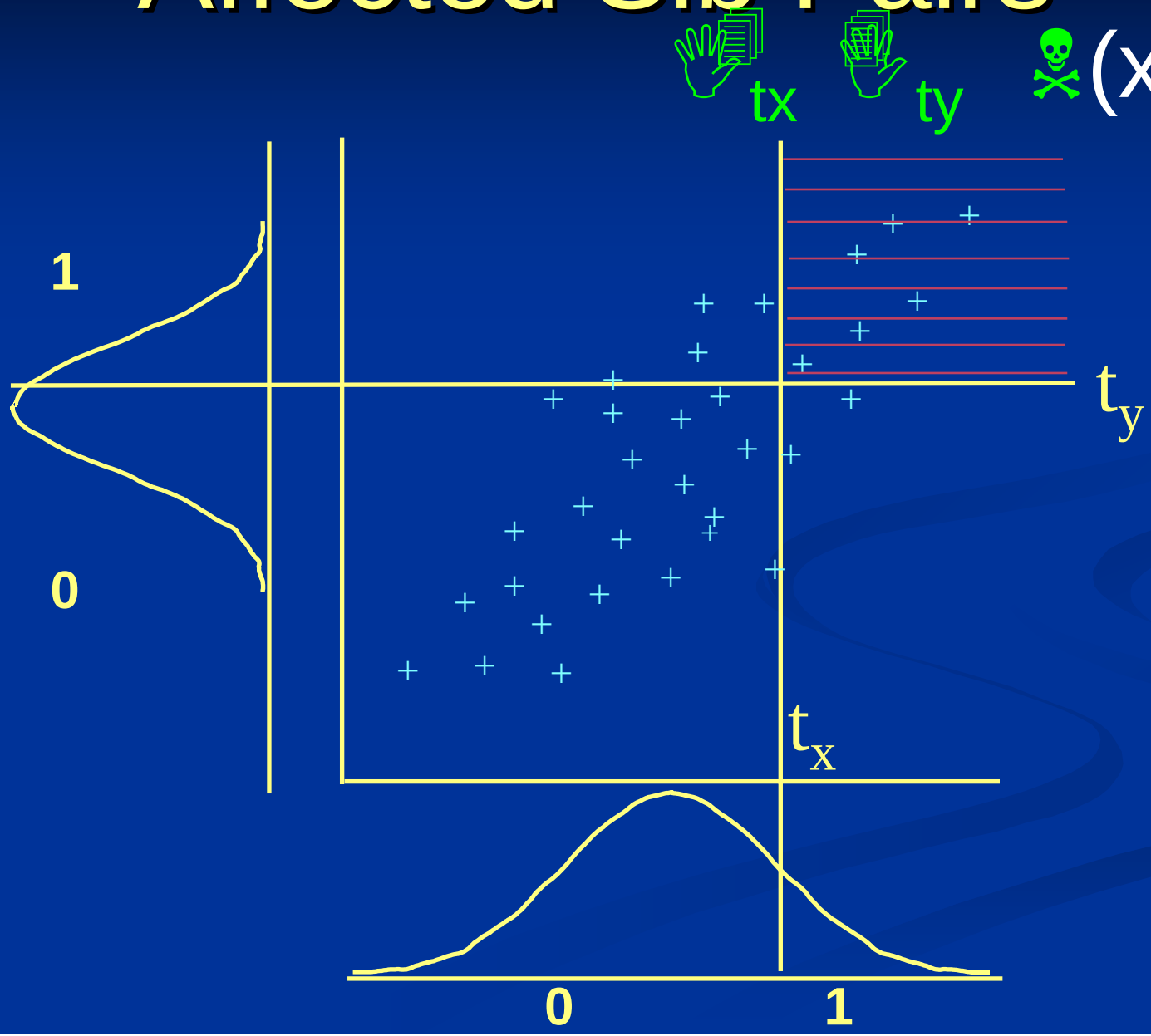


Ascertainment Correction

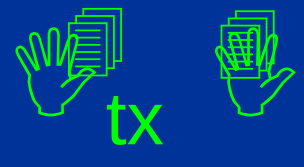
- Be / All you can be

$$\frac{\text{☠}(x)}{\int_{tx} \text{☠}(x)\,dx}$$

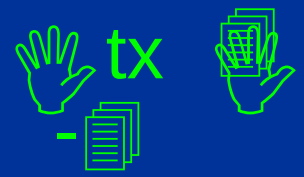
Affected Sib Pairs



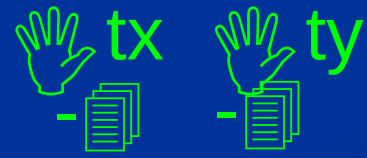
Ascertainment Corrections for Sib Pairs

ASP ++ 

☠(x,y) dy dx

DSP +- 

☠(x,y) dy dx

CUSP +- 

☠(x,y) dy dx

Correcting for ascertainment

Linkage studies

- Multivariate selection: multiple integrals
  - double integral for ASP
  - four double integrals for EDAC
- Use (or extend) weight formula
- Precompute in a calculation group
  - unless they vary by subject

Initial Results of Simulations

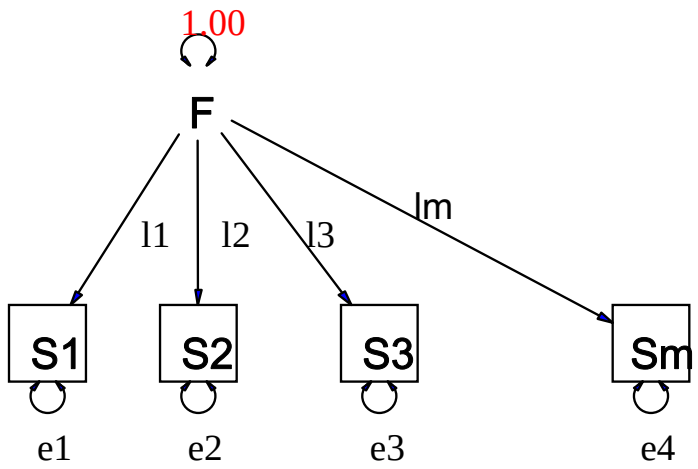
- Null Model
  - 50% heritability
  - No QTL
  - Used to generate null distribution
  - .05 empirical significance level at approximately 91 Chi-square
- QTL Simulations
  - 37.5% heritability
  - 12.5% QTL
  - Mx: 879 significant at nominal .05 p-value
  - Merlin: 556 significant at nominal .05 p-value
  - Some apparent increase in power

Measurement is KEY

Need continuous interval scales

Most complex traits not measured this way

Use latent trait instead



Factor model equivalent to Item response theory model

Can allow for non-normal Factors

Measurement of multiple Sx

Conclusion

- Quantifying QTL variation in selected samples can be done
- Can be computationally challenging
- May provide more power
- Permits multivariate analysis of correlated traits