Part I: Interval Mapping Basics

• observed measurements
  – \( Y \) = phenotypic trait
  – \( X \) = markers & linkage map
    • \( i \) = individual index 1,…,\( n \)
• missing data
  – missing marker data
  – \( Q \) = QT genotypes
    • alleles QQ, Qq, or qq at locus
• unknown quantities
  – \( \lambda \) = QT locus (or loci)
  – \( \theta \) = phenotype model parameters
• \( \text{pr}(Q|X,\lambda) \) recombination model
  – grounded by linkage map, experimental cross
  – recombination yields multinomial for \( Q \) given \( X \)
• \( \text{pr}(Y|Q,\theta) \) phenotype model
  – distribution shape (could be assumed normal)
  – unknown parameters \( \theta \) (could be non-parametric)

after Sen Churchill (2001)
recombination model components

\[ \lambda \]

distance along chromosome

\[ r_1 \quad r_2 \quad r_3 \quad r_4 \quad r_5 \]

recombination rates

\[ X_1 \quad X_2 \quad Q? \quad X_3 \quad X_4 \quad X_5 \quad X_6 \]

markers

recombination model components
Recombination and Distance

- assume map and marker distances are known
- useful approximation for QTL linkage
  - Haldane map function: no crossover interference
  - independence implies crossover events are Poisson
- all computations consistent in approximation
  - rely on given map with known marker locations
  - 1-to-1 relation of distance to recombination
  - all map functions are approximate anyway

\[
\begin{align*}
  r &= \frac{1}{2} \left( 1 - e^{-2\lambda} \right) \\
  \lambda &= -\frac{1}{2} \log(1 - 2r)
\end{align*}
\]
recombination model \( \text{pr}(Q|X, \lambda) \)

- locus \( \lambda \) is distance along linkage map
  - identifies flanking marker region
- flanking markers provide good approximation
  - map assumed known from earlier study
  - inaccuracy slight using only flanking markers
    - extend to next flanking markers if missing data
  - could consider more complicated relationship
    - but little change in results

\[
\text{pr}(Q|X, \lambda) = \text{pr}(\text{geno} | \text{map, locus}) \approx \text{pr}(\text{geno} | \text{flanking markers, locus})
\]
idealized phenotype model

- \text{trait} = \text{mean} + \text{additive} + \text{error}
- \text{trait} = \text{effect of geno} + \text{error}
- \text{pr} (\text{trait} | \text{geno, effects})

\[
Y = G_Q + E
\]

\[
\text{pr}(Y | Q, \theta) = \text{normal}(\hat{G}_Q, \sigma^2)
\]
Simulated Data with 1 QTL
Profile LOD for 1 QTL
What if data are far away from ideal?

- No QTL?
- Skewed?
- Dominance?
- Zeros?
What shape histograms by genotype?

WF/WF

WKy/WF

line = normal, + = semi-parametric, o = confidence interval
What QTL influence flowering time?
no vernalization: censored survival

- *Brassica napus*
  - Major female
    - needs vernalization
  - Stellar male
    - insensitive
  - 99 double haploids
- $Y = \log(\text{days to flower})$
  - over 50% Major at QTL never flowered
  - log not fully effective

grey = normal, red = non-parametric
What shape is flowering distribution?

*B. napus* Stellar

*B. napus* Major

line = normal, + = non-parametric, o = confidence interval
Who was Bayes?

• Reverend Thomas Bayes (1702-1761)
  – part-time mathematician
  – buried in Bunhill Cemetery, Moongate, London
  – famous paper in 1763 *Phil Trans Roy Soc London*
    • Barnard (1958 *Biometrika*), Press (1989) *Bayesian Statistics*
    • Stigler (1986) *History of Statistics*
    • Carlin Louis (1996); Gelman et al. (1995) books
  – Was Bayes the first with this idea? (Laplace)

• billiard balls on rectangular table
  – two balls tossed at random (uniform) on table
  – where is first ball if the second is to its right (left)?

first ball
Where is the first ball?

first ball

second ball

\[ \theta \]

prior
\[ \text{pr}(\theta) = 1 \]

likelihood
\[ \text{pr}(Y|\theta) = \theta^Y (1 - \theta)^{1-Y} \]

posterior
\[ \text{pr}(\theta|Y) = ? \]

\[ \text{pr}(\theta|Y) = \frac{\text{pr}(Y|\theta)\text{pr}(\theta)}{\text{pr}(Y)} \]

\[ \text{pr}(Y) = \int_0^1 \theta^Y (1 - \theta)^{1-Y} d\theta = \frac{1}{2} \]

\[ \text{pr}(\theta|Y) = \begin{cases} 
2\theta & Y = 1 \\
2(1 - \theta) & Y = 0
\end{cases} \]

(now throw second ball \( n \) times)
Likelihood and Posterior Example

data: \( Y = 1, 3, 8 \)

\[ \text{parameter: } \theta = ? \]

\[ \text{posterior} \]

\[ pr(Y = y | \theta) = \frac{\theta^y e^{-\theta}}{y!} \]

(M. Newton, pers. comm.)
effect of prior variance on posterior

\[ \kappa = 0.5 \]

\[ \kappa = 1 \]

\[ \kappa = 2 \]

normal prior, posterior for \( n = 1 \), posterior for \( n = 5 \), true mean
Bayesian Idea for QTLs

• key idea
  – sample missing genotypes $Q$
    • using recombination model
  – phenotype model given $Q$
    • see previous slides

• methods and philosophy
  – EM & MCMC
  – Frequentists & Bayesians

• review interval maps & profile LODs

• case study: simulated single QTL