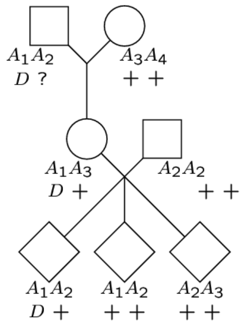
3.2 Likelihoods and Lod Scores (2 loci)
3.2.1 Counting Recombinants

- *Linkage analysis* is concerned with estimating ρ and with testing the . null hypothesis H0: $\rho = 1/2$, . against alternative H1: $\rho < 1/2$.
- Estimates and tests are based on likelihoods and likelihood ratios.
- In the figure: at a DNA marker locus, two grandparents have types A1A2 and A3A4; their daughter has type A1A3.
- She marries someone of type A2A2 and their three children are of types A1A2, A1A2 and A2A3.
- Granddad, the daughter, and the first child all carry some trait allele D. Other individuals carry only normal **+** alleles.

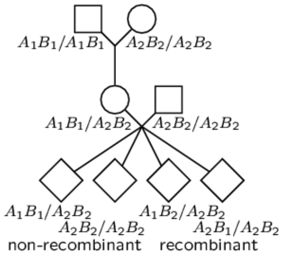


Example: continued

- The trait allele, D, segregates with the A1 marker allele from the grandad to his daughter, and the normal allele, **+**, segregates with A3 from grandma. That is, the grandparental data enable us to phase the mother.
- Note however that there is no information at all on grandparental phase. Hence no information about recombination from the grandparents to the mother.
- To the three children from their father, each receives an A2 **+** haplotype, regardless of recombination. This provides no information for linkage, but does enable us to identify haplotypes segregating from the mother to the children.
- To the three children from their mother, we have segregation of A1 with D, of A1 with **+**, and of A3 with **+**. Thus children 1 and 3 are non-recombinant (X1 = X3 =0) and child 2 is recombinant (X2 =1). So the number of scorable meioses n=3, and the number of recombinants T ~ Bin(3, ρ), and in this example T takes the value t=1.

3.2.2 Backcross Design (Phase known)

- Where each offspring can be classified recombinant or non-recombinant, as above, the number of recombinants in n observed offspring is T~ Bin (n, ρ).
- Such data arise in a *backcross experiment* using two inbred lines.
- Line 1: alleles A1 and B1 (genotype A1B1/ A1B1). Line 2: alleles A2 and B2 (genotype A2B2/ A2B2).
- Hybrid (F1): all have genotype A1B1/ A2B2.
- Backcross to line 2: all offspring get A2B2 from the line-2 parent; combination A1B1, or A2B2 (non-recombinant), or A1B2 or A2B1 (recombinant) from the F1 parent observable.



Backcross ctd: MLE of ρ and Lod scores

- Suppose n offspring of such matings are scored, and a total t are recombinant. It does not matter whether these are in the same of different matings, since all are independent.
- T ~ Bin(n, ρ) so $\lambda(\rho) = t \log(\rho) + (n – t) \log(1- \rho)$.
- To test for linkage, compare the likelihood to its value in the absence of linkage (ρ= 1/2): the log-likelihood difference is . $\text{lod}(\rho) = \lambda(\rho) – \lambda(1/2) = t \log(\rho) + (n – t) \log(1- \rho) + n \log (2)$.
- With base-10 logs, this is known as the *lod score.*
- The MLE of ρ is $\rho^*= t/n$, provided $2t \le n$ (since $\rho \le 1/2$).
- To test $\rho = 1/2$ against $\rho < 1/2$, the maximized lod score is: . $\text{lod}(\rho^*) = t \log t + (n-t) \log (n-t) - n \log(n/2)$. provided $2t \le n$, and 0 otherwise.
- This is a decreasing function of t, and we reject the null hypothesis ρ =1/2 if t < t0 with critical value t0 chosen to give a specified size of the test (type I error).

3.2.3 Type-1 Error and Critical Values

- When n is large, T is approximately $N(n\,\rho,\ n\,\rho(1-\rho))$.
- If $\rho = ½$, $T \sim N(n/2,\ n/4)$, is a very good approximation.
- Then $(2/\sqrt{n})(T - n/2) \sim N(0,1)$.
- For a test size α, reject $H0: \rho=1/2$ in favor of $H1: \rho \le 1/2$ if . $(2/\sqrt{n})(T - n/2) \le \Phi^{-1}(\alpha)$ where $\Phi(.)$ is the standard Normal cdf.
- For example, for $\alpha = 0.025$, . $\Phi^{-1}(\alpha) = -1.96 \approx -2$, . so reject H0 if $T \le n/2 - \sqrt{n} = t0$.
- The table shows critical values for a test size $\alpha = 0.025$ and corresponding base-10 lod scores for binomial samples.
- Also shown is t0 required to give a lod score of 3.

n	≈ t0	≈ t0/n	lod at (t0/n)	t0 for lod 3
25	7	0.3	1.088	≤ 3
100	40	0.4	0.874	≤ 31
625	287	0.46	0.905	≤ 267
1024	480	0.48	0.869	≤ 452

Prior probability of linkage

- The (base 10) lod score is around 1 for a number of recombinants at the critical value for a test of size $\alpha = 0.025$ of $H0: \rho = 1/2$.
- Traditionally, a base-10 lod score of 3 is required to infer linkage. We see from the table that this is a more stringent test. For example, if n=100, we will reject $H0: \rho = 1/2$ with type 1 error $\alpha = 0.025$ if t is less than t0=40, but for a lod score of 3 we would need t less than 31. This is a type 1 error or about 0.0001.
- The idea was that if two arbitrary locations in the genome are chosen the prior probability of linkage is small, about 0.05, so that strong evidence is needed to reject H0.
- Nowadays, with genome-wide scans this is not so relevant. Instead, we have a multiple testing problem. However, the convention of a base-10 lod score of 3 still stands.
- For markers, and simple Mendelian traits, few if any lod scores of 3 or more have been subsequently found to be false positives, whereas quite a few between 2 and 3 have been later shown to be false.

3.2.4 TESTING USING LOD SCORES

- We can use the (base e) lod score in a likelihood-based test for linkage: $\lambda(\rho) = t \log(\rho) + (n-t)\log(1-\rho)$ and the lod score is . $\mathrm{lod}(\rho) = \lambda(\rho) - \lambda(1/2)$.
- The MLE is $\rho^* = t/n$ (assuming this is ≤ 0.5) and . $\lambda(\rho^*) = t \log t + (n-t)\log(n-t) - n \log n$
- Example 1: $H0: \rho = 0.1$.
  Then $2(\lambda(\rho^*) - \lambda(0.1)) \sim \chi^2_1$ if H0 is true.
- Example 2: But to test for linkage, we want $H0: \rho = 0.5$.
  Then $2(\lambda(\rho^*) - \lambda(0.5)) = 2\,\mathrm{lod}(\rho^*)$.
  If H0 is true, then half the time $\rho^* = 0.5$, and $\lambda(\rho^*) = \lambda(0.5)$. So then $2(\lambda(\rho^*) - \lambda(0.5))$ is $(1/2) \times 0 + (1/2) \times \chi^2_1$ if there is no linkage.
- This means that for this case $4(\lambda(\rho^*) - \lambda(0.5))$ is χ^2_1 if H0 is true. This extra factor of 2 can be confusing: for the counting recombinants case it may be simpler to stick to the Normal test.
- Note in fact this corresponds to the fact that in our test we did a one-sided test; $\alpha = 0.025$ at 1.96 or ≈2 st.dev, instead of $\alpha = 0.05$.

Sex-specific recombination rates

- Example 3: Suppose we see t_m recombinants in n_m male meioses and t_f recombinants in n_f female meioses. Then we can test $H0: \rho_m = \rho_f$.
- Unconstrained case (general hypothesis):
  $\lambda(\rho_m, \rho_f) = t_m \log(\rho_m) + (n_m - t_m)\log(1-\rho_m)$. $+ t_f \log(\rho_f) + (n_f - t_f)\log(1-\rho_f)$ maximized by $\rho^*_m = t_m / n_m$, $\rho^*_f = t_f / n_f$.
- Under $H0: \rho_m = \rho_f = \rho$ say. Then $\lambda(\rho, \rho)$ is $(t_m + t_f)\log(\rho) + ((n_m+n_f) - (t_m+t_f))\log(1-\rho)$ maximized by $\rho^* = (t_m+t_f)/(n_m+n_f)$.
- If H0 is true, $2(\lambda(\rho^*_m, \rho^*_f) - \lambda(\rho^*, \rho^*))$ is χ^2_1.
- This is a 1 degree of freedom test as there are 2 parameters in general (i.e. ρ_m, ρ_f), and 1 (i.e. ρ) under H0.