

Two-dimensional, two-QTL scans

Most complex traits are understood to be the result of the action of multiple genetic loci. Some loci may be linked, and the effect of some loci may depend on the genotype at others (epistasis). In this chapter, we undertake the first step in the direction of teasing apart the multiple linked or interacting QTL underlying a quantitative trait, by contemplating two-dimensional, two-QTL genome scans.

Thus far we have focused on one-dimensional genome scans, in which one fits all possible single-QTL models. Despite the fact that such genome scans appear to ignore the reality of complex traits (with multiple loci, possibly linked or epistatic), they have worked quite well. Part of this success is attributable to the fact that even if multiple loci are contributing to a trait, they are often on separate chromosomes and have marginal effects large enough to be detectable. Because of the independent assortment of chromosomes (Mendel's second law, that genotypes at loci on different chromosomes are independent), genome scans treating the chromosomes separately give approximately the same results as would be obtained if one had conditioned on additive QTL with small effect on other chromosomes. In this sense, one-dimensional genome scans transcend their one-dimensional nature.

Nevertheless, there are some limitations to adopting a one-dimensional approach to the essentially multidimensional problem of detecting multiple QTL. The joint consideration of multiple QTL has three advantages. First, by taking account of QTL of large effect, one may reduce the residual variation and so better identify QTL of modest effect. Second, the separation of linked QTL is best achieved by comparing the fit of a two-QTL model to the best single-QTL model. (With two distinct peaks in the LOD curve for a given chromosome, one may be tempted to infer the presence of two QTL, and there is some literature considering the shape of LOD curves to infer the presence of multiple linked loci, an effort that we term the *phrenology* of LOD curves. It is far better to simply assess the improvement in fit that comes from a two-QTL model.) Finally, epistasis between QTL may only be assessed via models that explicitly consider multiple QTL.

The obvious next step, beyond a one-dimensional genome scan, is consideration of all possible two-locus QTL models in a two-dimensional genome scan. This allows us to identify potential interactions among QTL and to assess evidence for linked QTL. The interpretation of the results of two-dimensional genome scans is less straightforward than those of one-dimensional scans. That is the subject of this chapter.

We begin by considering two-dimensional genome scans in the context of normally distributed traits. We then discuss the analysis of binary traits. The special treatment of the X chromosome is covered briefly. Finally, we discuss methods to handle covariates.

8.1 The normal model

Imagine the presence of precisely two QTL, and consider each pair of positions in the genome as the putative locations for those QTL. Let (s, t) denote the pair of positions, and consider the following four models; for now we assume that the residual variation is normally distributed with constant variance.

$$\begin{aligned} H_f: y &= \mu + \beta_1 q_1 + \beta_2 q_2 + \gamma(q_1 \times q_2) + \epsilon \\ H_a: y &= \mu + \beta_1 q_1 + \beta_2 q_2 + \epsilon \\ H_1: y &= \mu + \beta_1 q_1 + \epsilon \\ H_0: y &= \mu + \epsilon \end{aligned} \tag{8.1}$$

In the full model (H_f), the two QTL are allowed to interact. In the additive model (H_a), they are assumed to act additively (i.e., the effect of each locus is the same, no matter the genotype at the other locus). We also consider all possible single-QTL models (i.e., the results of the single-QTL genome scan), and the null model (H_0), that there are no QTL.

The fit of these models requires that we take account of the missing genotype data at the putative QTL, making use of the genotype data at linked markers. This may again be accomplished in multiple ways; the four methods discussed in Chap. 4 (maximum likelihood via the EM algorithm, Haley–Knott regression, the extended Haley–Knott method, and multiple imputation) may all be readily extended to the two-QTL case. In the case that the two putative QTL are on the same chromosome, the multiple imputation approach requires that genotypes at the two positions be drawn from their joint distribution, but this is easily done (see Sec. D.3). The other methods require the calculation of the joint genotype probabilities, $\Pr(q_1, q_2 | \mathbf{M})$, where $\mathbf{M}$ is the observed multipoint marker genotype data. If one assumes no crossover interference and no genotyping errors, and if there is complete genotype data at an intervening marker, then the QTL genotypes will be conditionally independent, given the marker data, and so

$$\Pr(q_1, q_2 | \mathbf{M}) = \Pr(q_1 | \mathbf{M}) \times \Pr(q_2 | \mathbf{M}). \tag{8.2}$$

If we wish to allow for genotyping errors, or if there is no fully typed marker between the putative QTL, then the product rule in equation (8.2) will not apply. However, with the assumption of no crossover interference, the joint distribution may be calculated efficiently via the hidden Markov model technology (see Sec. D.4).

Turning now to the summary and interpretation of two-dimensional, two-QTL genome scans, let $l_f(s, t)$ denote the $\log_{10}$ likelihood for the full model with QTL at s and t , $l_a(s, t)$ denote the $\log_{10}$ likelihood for the additive model with QTL at s and t , $l_1(s)$ denote the $\log_{10}$ likelihood for the single-QTL model with the QTL at s , and l_0 denote the $\log_{10}$ likelihood under the null (with no QTL).

We may immediately define several LOD scores, comparing the fit of the four models.

$$\begin{aligned}\text{LOD}_f(s, t) &= l_f(s, t) - l_0 \\ \text{LOD}_a(s, t) &= l_a(s, t) - l_0 \\ \text{LOD}_i(s, t) &= l_f(s, t) - l_a(s, t) \\ \text{LOD}_1(s) &= l_1(s) - l_0\end{aligned}$$

LOD_f measures the improvement in the fit of the full two-locus model over the null model, and indicates evidence for at least one QTL, with allowance for interaction. Similarly, LOD_a measures the improvement in the fit of the two-locus additive model, and indicates evidence for at least one QTL, assuming no interaction. LOD_i measures the improvement in the fit of the full model over that of the additive model, and so indicates evidence for interaction. LOD_1 is simply the LOD score from the one-dimensional, single-QTL genome scan.

These LOD scores can be difficult to interpret, as LOD_f and LOD_a will be large on any chromosome with clear evidence for a QTL. That is, if the initial genome scan indicated strong evidence for a QTL at position s_0 , so that $\text{LOD}_1(s_0)$ is large, then $\text{LOD}_f(s_0, t)$ and $\text{LOD}_a(s_0, t)$ will be large for all t .

Most important, in the consideration of the results of a two-dimensional, two-QTL genome scan, is an assessment of evidence for a second QTL: does the two-QTL model provide sufficiently improved fit over the best single-QTL model? We find it valuable to focus on an individual chromosome or pair of chromosomes. Consider a pair of chromosomes (j, k) , including the case $j = k$, and let $c(s)$ denote the chromosome for position s . We now consider the maximum LOD scores over that pair of chromosomes.

$$\begin{aligned}M_f(j, k) &= \max_{c(s)=j, c(t)=k} \text{LOD}_f(s, t) \\ M_a(j, k) &= \max_{c(s)=j, c(t)=k} \text{LOD}_a(s, t) \\ M_1(j, k) &= \max_{c(s)=j \text{ or } k} \text{LOD}_1(s)\end{aligned}$$

So $M_f(j, k)$ is the $\log_{10}$ likelihood ratio comparing the full model with QTL on chromosomes j and k to the null model, and $M_a(j, k)$ is the analogous thing for the additive model. Note that the pair of positions at which the full model is maximized may be different from the pair of positions at which the additive model is maximized. $M_1(j, k)$ is the $\log_{10}$ likelihood ratio comparing the model with a single QTL on either chromosomes j or k to the null model.

We derive three further LOD scores from the above.

$$\begin{aligned} M_i(j, k) &= M_f(j, k) - M_a(j, k) \\ M_{fv1}(j, k) &= M_f(j, k) - M_1(j, k) \\ M_{av1}(j, k) &= M_a(j, k) - M_1(j, k) \end{aligned}$$

$M_i(j, k)$ is the $\log_{10}$ likelihood ratio comparing the full model with QTL on chromosomes j and k to the additive model with QTL on chromosomes j and k , and so indicates evidence for an interaction between QTL on chromosomes j and k , assuming that there is precisely one QTL on each chromosome (or, for $j = k$, that there are two QTL on the chromosome).

$M_{fv1}(j, k)$ is the $\log_{10}$ likelihood ratio comparing the full model with QTL on chromosomes j and k to the single-QTL model, with a single QTL on either chromosome j or k . Thus, it indicates evidence for a second QTL, allowing for the possibility of epistasis.

$M_{av1}(j, k)$ is the $\log_{10}$ likelihood ratio comparing the additive model with QTL on chromosomes j and k to the single-QTL model, with a single QTL on either chromosome j or k . Thus, it indicates evidence for a second QTL, assuming no epistasis.

We focus particularly on M_{fv1} and M_{av1} , concerning evidence for a second QTL. M_{fv1} allows for interactions between the QTL, and so will enable the detection of loci with limited marginal effect. However, in the absence of a strong interaction, M_{av1} would give greater power to detect a second QTL, as the additional degrees of freedom in the M_{fv1} statistic results in a greater threshold for significance.

In order to identify loci of interest from the results of a two-dimensional, two-QTL genome scan, we consider the distributions of the five LOD scores ($M_f(j, k)$, $M_{fv1}(j, k)$, $M_i(j, k)$, $M_a(j, k)$ and $M_{av1}(j, k)$) under the global null hypothesis, that there are no QTL. Through either computer simulation or a permutation test, we may estimate quantiles of the null distributions of these LOD scores, to be used as thresholds. We use the following rule: A pair of chromosomes (j, k) is reported as interesting if either of the following holds:

$$\begin{cases} M_f(j, k) \geq T_f \text{ and } [M_{fv1}(j, k) \geq T_{fv1} \text{ or } M_i(j, k) \geq T_i] \\ M_a(j, k) \geq T_a \text{ and } M_{av1}(j, k) \geq T_{av1} \end{cases} \quad (8.3)$$

We are inclined to ignore $M_i(j, k)$ in this rule (i.e., setting $T_i = \infty$) and use a common significance level ($\alpha = 5$ or 10%) for the four remaining thresholds.

The thresholds can be obtained by a permutation test (see below), but this is extremely time-consuming. For a mouse backcross, we suggest the 5%

thresholds $(T_f, T_{fv1}, T_i, T_a, T_{av1}) = (6.0, 4.7, 4.4, 4.7, 2.6)$. For a mouse intercross, we suggest $(T_f, T_{fv1}, T_i, T_a, T_{av1}) = (9.1, 7.1, 6.3, 6.3, 3.3)$. These are the estimated 95th percentiles of the null distributions of the corresponding LOD scores, obtained by 10,000 simulations of crosses with 250 individuals, markers at a 10 cM spacing, and analysis by Haley–Knott regression.

While we are recommending that the thresholds be derived under the global null hypothesis of no QTL, our interest is really in evidence for a second QTL, over and above that for a single QTL. Thus, we really should be considering the distributions of these statistics in the presence of a single QTL: how large will M_{fv1} and M_{av1} be, if there exists precisely one QTL? However, this is not a simple matter, and the answer would likely depend, to some extent, on the location and effect of the hypothesized QTL. And so we make the assumption that the distributions of M_{fv1} and M_{av1} are approximately the same, whether there exists a single QTL or no QTL.

Our separate treatment of individual pairs of chromosomes may be confusing initially. Our goal, with this approach, is to identify multiple pairs of linked or interacting QTL across the genome, in the same way that one seeks, in a traditional one-dimensional genome scan, to identify QTL on multiple chromosomes, rather than just the single locus giving the largest LOD score. Our test statistics M_{fv1} and M_{av1} are equivalent to the usual sort of likelihood ratio statistics for comparing a model with two QTL to a model with a single QTL.

Example

As an illustration, we return to the `hyper` data (see Sec. 2.3). Let us load the data and run `calc.genoprob` to calculate the conditional QTL genotype probabilities, given the available marker data. We use a more coarse grid, `step=2.5`, for the sake of computational speed.

```
> library(qt1)
> data(hyper)
> hyper <- calc.genoprob(hyper, step=2.5, err=0.001)
```

The two-dimensional, two-QTL scan is accomplished with the function `scantwo`, which operates much like `scanone`, though computation time is greatly increased. By default, the analysis is performed by maximum likelihood via the EM algorithm (`method="em"`). We use `verbose=FALSE` to suppress tracing information.

```
> out2 <- scantwo(hyper, verbose=FALSE)
```

The output has class `"scantwo"`, and so may be plotted with `plot.scantwo` and summarized with `summary.scantwo`. Each of these functions includes considerable flexibility. First, let us plot the two-dimensional scan results for selected chromosomes.

```
> plot(out2, chr=c(1,4,6,7,15))
```

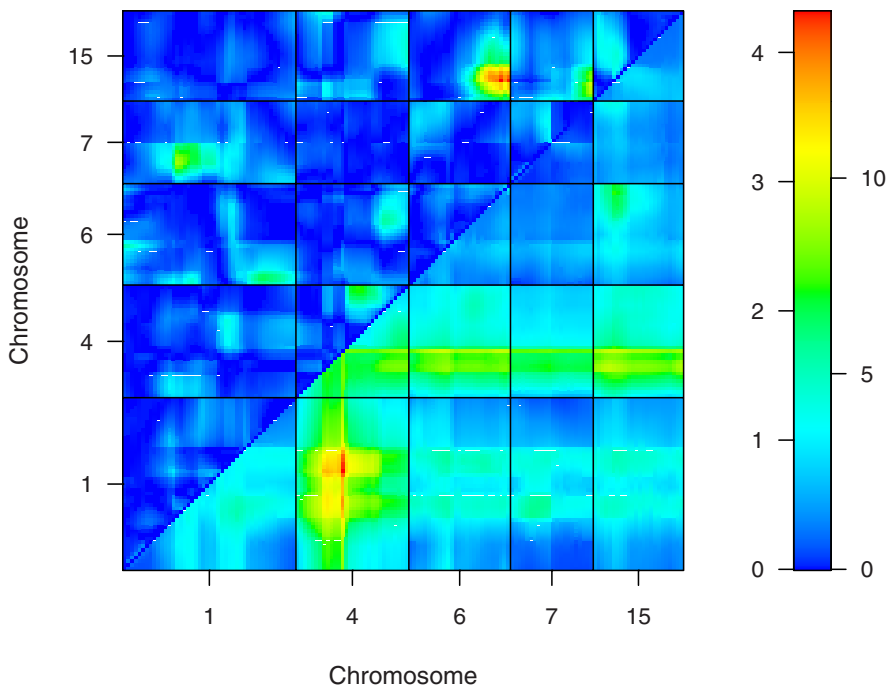


Figure 8.1. LOD scores, for selected chromosomes, from a two-dimensional, two-QTL genome scan with the **hyper** data. LOD_i is displayed in the upper left triangle; LOD_f is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_f , respectively.

The result is shown in Fig. 8.1. By default, LOD_i is plotted in the upper left triangle, and LOD_f is plotted in the lower right triangle. In the color scale on the right, the numbers on the left and right correspond to LOD_i and LOD_f , respectively. Note that LOD_f (lower right triangle) is large for chromosome 4 considered with any other chromosome. These “tails” in LOD_f are due to the strong evidence for a QTL on chromosome 4, and the fact that LOD_f compares the full two-QTL model to the null model, and so is large in the presence of evidence for at least one QTL. Further note the evidence for an interaction between loci on chromosomes 6 and 15. (LOD_i , in the upper left triangle, is large.)

In place of LOD_f , we prefer to focus on LOD_{fv1} , which indicates evidence for a second QTL, allowing for epistasis. Above, we had defined $M_{fv1}(j, k) = M_f(j, k) - M_1(j, k)$, for a pair of chromosomes (j, k) . We now consider $\text{LOD}_{fv1}(s, t) = \text{LOD}_f(s, t) - M_1(j, k)$, though replacing negative values with 0. $\text{LOD}_{av1}(s, t)$ may be defined similarly.

We may plot LOD_{fv1} in the lower right triangle, in place of LOD_f , with the argument `lower="cond-int"`, as follows. (One may also use `lower="fv1"`.)

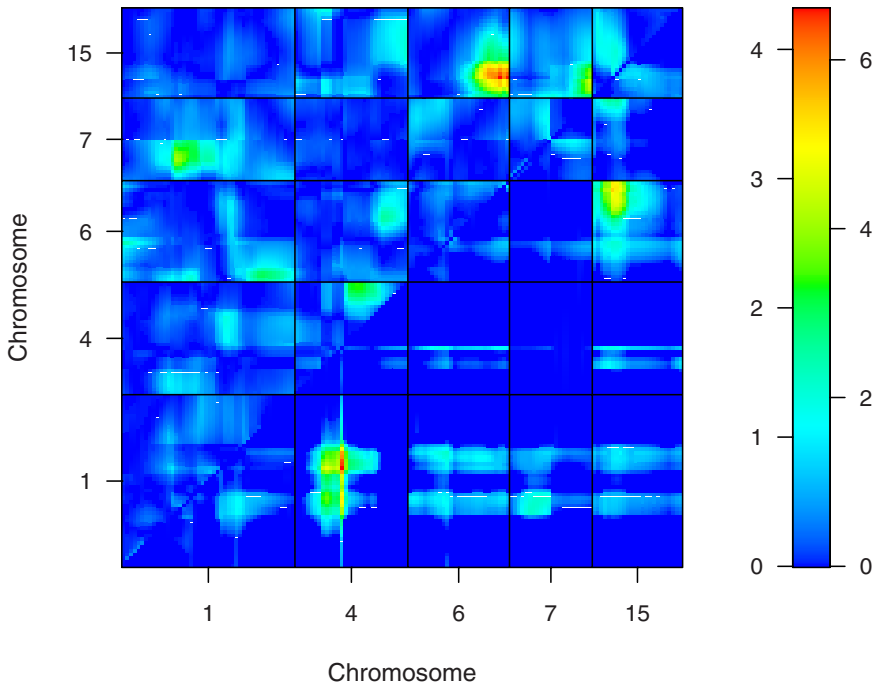


Figure 8.2. LOD scores, for selected chromosomes, from a two-dimensional, two-QTL genome scan with the **hyper** data. LOD_i is displayed in the upper left triangle; LOD_{fv1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_{fv1} , respectively.

```
> plot(out2, chr=c(1,4,6,7,15), lower="cond-int")
```

In the plot of LOD_{fv1} , in Fig. 8.2, the lower right triangle is cleaned up considerably. We now see strong evidence for a pair of QTL on chromosomes 1 and 4, and on chromosomes 6 and 15. (That is, for these pairs, the fit of the two-QTL model, with one QTL on each chromosome, considerably improves on the fit of the best single-QTL model, with a single QTL on one or the other chromosome.) The tails that came from the chromosome 4 locus have been eliminated, making the figure more easily interpreted.

One may also plot LOD_a and/or LOD_{av1} , and what is plotted in the upper triangle may be modified in the same way as we modified what was plotted in the lower right triangle. For example, we may plot LOD_a and LOD_{av1} as follows. To get LOD_{av1} in the lower right triangle, we may either use `lower="cond-add"` or `lower="av1"`.

```
> plot(out2, chr=c(1,4,6,7,15), upper="add", lower="cond-add")
```

The result is displayed in Fig. 8.3. Note that LOD_a , like LOD_f , has the same “tails” problem (with large LOD scores appearing on chromosome 4

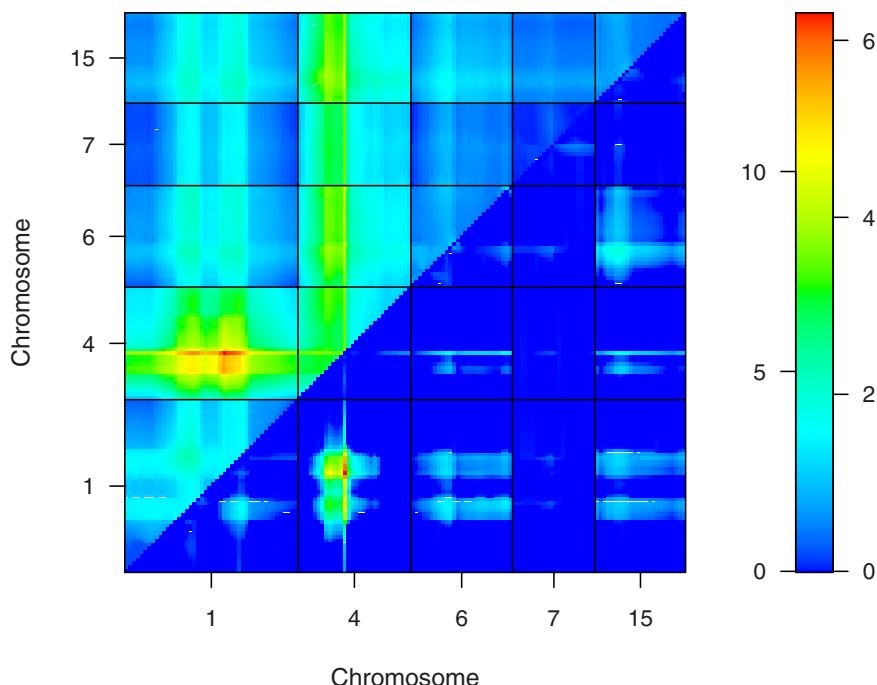


Figure 8.3. LOD scores, for selected chromosomes, from a two-dimensional, two-QTL genome scan with the **hyper** data. LOD_a is displayed in the upper left triangle; LOD_{av1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_a and LOD_{av1} , respectively.

together with any other chromosome). Also note that, for chromosomes 6 and 15, the evidence for a second QTL has disappeared. These loci are seen as important only when they are allowed to interact.

Recall the two peaks in the LOD curve for chromosome 1 with the **hyper** data. (For example, see Fig. 4.6 on page 88.) We are thus particularly interested in evidence for a second QTL on chromosome 1. We may plot LOD_{av1} and LOD_{fv1} for chromosome 1 as follows.

```
> plot(out2, chr=1, lower="cond-int", upper="cond-add")
```

The results, in Fig. 8.4, indicate little evidence for a second QTL on chromosome 1. Inclusion of a second locus increases the $\log_{10}$ likelihood by ~ 1.5 .

The plots of the results of the two-dimensional, two-QTL scan are not so simple to interpret as those of a one-dimensional, single-QTL scan. Thus, we place greater reliance on the numeric summaries from **summary.scantwo**. Use of **summary(out2)** would produce a table giving a single row for each pair of chromosomes. With 20 chromosomes, that is a table with 210 rows. That is an unwieldy amount of information to sift through, and so it is best to provide a

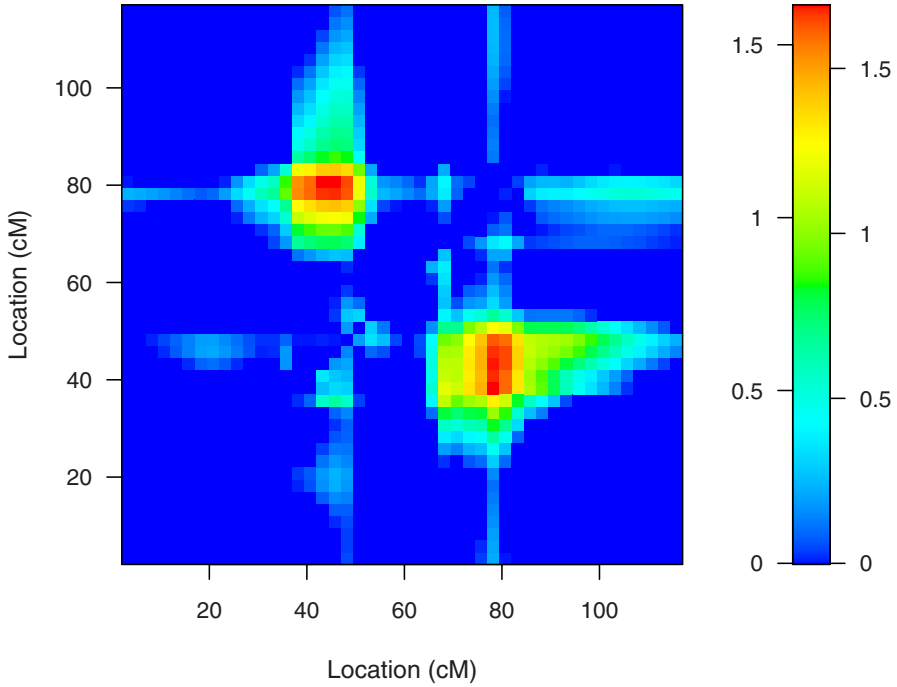


Figure 8.4. LOD scores, for chromosome 1, from a two-dimensional, two-QTL genome scan with the **hyper** data. LOD_{av1} is displayed in the upper left triangle; LOD_{fv1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_{av1} and LOD_{fv1} , respectively.

set of thresholds, $(T_f, T_{fv1}, T_i, T_a, T_{av1})$; then, only those chromosomes satisfying the rule in equation (8.3) (page 216) will be displayed.

For example, we may use the thresholds $(T_f, T_{fv1}, T_i, T_a, T_{av1}) = (6.0, 4.7, 4.4, 4.7, 2.6)$, suggested by simulations.

```
> summary(out2, thresholds=c(6.0, 4.7, 4.4, 4.7, 2.6))
```

	pos1f	pos2f	lod.full	lod.fv1	lod.int	pos1a	pos2a
c1:c4	68.3	30	14.24	6.6	0.305	68.3	30.0
c6:c15	60.0	18	7.27	5.4	3.458	25.0	20.5
	lod.add		lod.av1				
c1:c4	13.94	6.30					
c6:c15	3.81	1.95					

There is clear evidence for a pair of QTL on chromosomes 1 and 4, and this is true whether or not an interaction is allowed ($M_{fv1} = 6.6$ and $M_{av1} = 6.3$). We see little evidence for an interaction between these loci ($M_i = 0.3$). Note that (**pos1f**, **pos2f**) indicate the estimated positions of the QTL (on chromosomes 1 and 4, respectively) under the full model, while (**pos1a**, **pos2a**)

are the estimated positions under the additive model. These are allowed to be different, though for the chromosomes (1,4) pair, the likelihoods for the full and additive models happened to be maximized at the same pair of positions.

The summary also indicates strong evidence for the pair of QTL on chromosomes 6 and 15, though only if an interaction is allowed ($M_{fv1} = 5.4$ and $M_{av1} = 1.9$), and here the full model was maximized at a different pair of positions than the additive model. In the summary, the evidence for interaction, $M_i = 3.5$ is the difference $M_f - M_a$, or equivalently $M_{fv1} - M_{av1}$, with the full and additive models allowed to be maximized at different positions.

As mentioned above, we are inclined to ignore the value of M_i , by setting $T_i = \infty$. This is accomplished as follows, though here the same summary is obtained.

```
> summary(out2, thresholds=c(6.0, 4.7, Inf, 4.7, 2.6))
```

	pos1f	pos2f	lod.full	lod.fv1	lod.int		pos1a	pos2a
c1:c4	68.3	30	14.24	6.6	0.305		68.3	30.0
c6:c15	60.0	18	7.27	5.4	3.458		25.0	20.5

	lod.add	lod.av1
c1:c4	13.94	6.30
c6:c15	3.81	1.95

The thresholds above were obtained by simulation of a backcross with a genome modelled after that of the mouse, markers at a 10 cM spacing, and with Haley–Knott regression used for the analysis. While these may serve as a reasonably general guide, it would be better to use a permutation test with the observed data, so that the results would take account of the phenotype distribution, marker density, and pattern of missing genotype data. A permutation test may be accomplished with `scantwo` by specifying the argument `n.perm`, though the required computation time may be daunting.

Because a selective genotyping strategy was used for the `hyper` data, it is best to use a stratified permutation test, splitting the individuals into two strata according to the amount of genotype data available, and permuting the phenotypes relative to the genotypes separately within the strata. This may be accomplished by creating a vector indicating the two strata, to be specified via the `perm.strata` argument in `scantwo`. Thus, we perform the permutation test for the two-dimensional, two-QTL genome scan as follows.

```
> strat <- (nmissing(hyper) > 50)
> operm2 <- scantwo(hyper, n.perm=1000, perm.strat=strat)
```

The permutation results have class `"scantwoperm"`. We may use `summary.scantwoperm` to get estimated thresholds.

```
> summary(operm2)

bp (1000 permutations)
```

```

      full  fv1  int  add  av1  one
5%  6.13 4.86 4.34 4.64 3.04 2.82
10% 5.63 4.48 3.95 4.27 2.73 2.48

```

The 5% thresholds are similar to those presented previously, though T_{av1} is a bit larger.

Because the computation time for the permutations can be on the order of 100 hours, one may want to split the calculations across multiple computers. In doing so, one should be careful about the seed for the random number generator. This is saved as the object `.Random.seed` in the R workspace. If multiple sets of permutations are run in parallel from the same R workspace, one may unintentionally use the same seed, and so obtain identical results, for each set of permutations. Thus, one should call `set.seed` separately for each set.

The permutations above could be run in two batches using the following pieces of code.

```

> set.seed(85842518)
> operm2a <- scantwo(hyper, n.perm=500, perm.strat=strat)
> save(operm2a, file="perm2a.RData")

> set.seed(85842519)
> operm2b <- scantwo(hyper, n.perm=500, perm.strat=strat)
> save(operm2b, file="perm2b.RData")

```

The two bits can then be combined with the `c.scantwoperm` function.

```

> load("perm2a.RData")
> load("perm2b.RData")
> operm2 <- c(operm2a, operm2b)

```

The same technique may also be used for a permutation test in a one-dimensional genome scan, and there is a function `c.scanoneperm` for combining multiple batches of permutation replicates.

We may include the permutation results in the call to `summary.scantwo`, so that the thresholds are calculated automatically, and so that p -values may be calculated. In place of thresholds, we must provide significance levels. We may ignore the M_i values by giving a significance level of 0 (which corresponds to $T_i = \infty$). Using 20% significance levels, we obtain the following.

```

> summary(out2, perms=operm2, alphas=c(0.2, 0.2, 0, 0.2, 0.2),
+         pvalues=TRUE)

```

	pos1f	pos2f	lod.full	pval	lod.fv1	pval	lod.int	pval
c1:c4	68.3	30.0	14.24	0.000	6.60	0.003	0.305	1.00
c3:c3	37.2	44.7	4.53	0.493	3.75	0.348	0.215	1.00
c6:c15	60.0	18.0	7.27	0.013	5.40	0.023	3.458	0.25

	pos1a	pos2a	lod.add	pval	lod.av1	pval
c1:c4	68.3	30.0	13.94	0.000	6.30	0.000
c3:c3	37.2	44.7	4.32	0.092	3.53	0.011
c6:c15	25.0	20.5	3.81	0.205	1.95	0.444

Note that, if we had wanted to consider a threshold on the interaction LOD score, M_i , we would have instead typed `alphas=rep(0.2,5)`, or, as short hand, simply `alphas=0.2`.

In addition to the loci on chromosomes 1, 4, 6 and 15, we see evidence for a pair of linked, additive QTL on chromosome 3. But these are very tightly linked (separated by just 7.5 cM), and so may be an artifact.

Two-dimensional, two-QTL scans often show artifacts along the diagonal (that is, for two tightly linked QTL), particularly in cases in which two putative loci are not separated by a typed marker. One thus may find value in the utility function `clean.scantwo`, which replaces LOD scores, for pairs of positions that are not separated by at least one marker, with 0, and so eliminates the bulk of such artifacts. This will be done automatically within `scantwo` if one uses the argument `clean.output=TRUE`, and this is particularly recommended for permutations.

We may plot the two-dimensional scan results for chromosome 3, zeroing out the LOD scores for pairs of positions that are not separated by a marker, as follows.

```
> plot(clean(out2), chr=3, lower="cond-add")
```

In the results, in Fig. 8.5, large blue squares along the diagonal correspond to pairs of positions that are not separated by a marker. Note that the peak in the LOD surface remains. This may also be seen in the summary of the “cleaned” results:

```
> summary(clean(out2), perms=operm2,
+         alphas=c(0.2, 0.2, 0, 0.2, 0.2))
```

	pos1f	pos2f	lod.full	lod.fv1	lod.int	pos1a	pos2a
c1:c4	68.3	30.0	14.24	6.60	0.305	68.3	30.0
c3:c3	37.2	44.7	4.53	3.75	0.215	37.2	44.7
c6:c15	60.0	18.0	7.27	5.40	3.458	25.0	20.5

	lod.add	lod.av1
c1:c4	13.94	6.30
c3:c3	4.32	3.53
c6:c15	3.81	1.95

Let us study the effects of the putative linked loci on chr 3 via the functions `plot.pxg` and `effectplot`. The `effectplot` function requires imputed genotype data, and so we first run `sim.geno`, though only for chromosome 3.

```
> hyperc3 <- sim.geno(subset(hyper, chr=3), step=2.5,
+                     error.prob=0.001, n.draws=256)
```

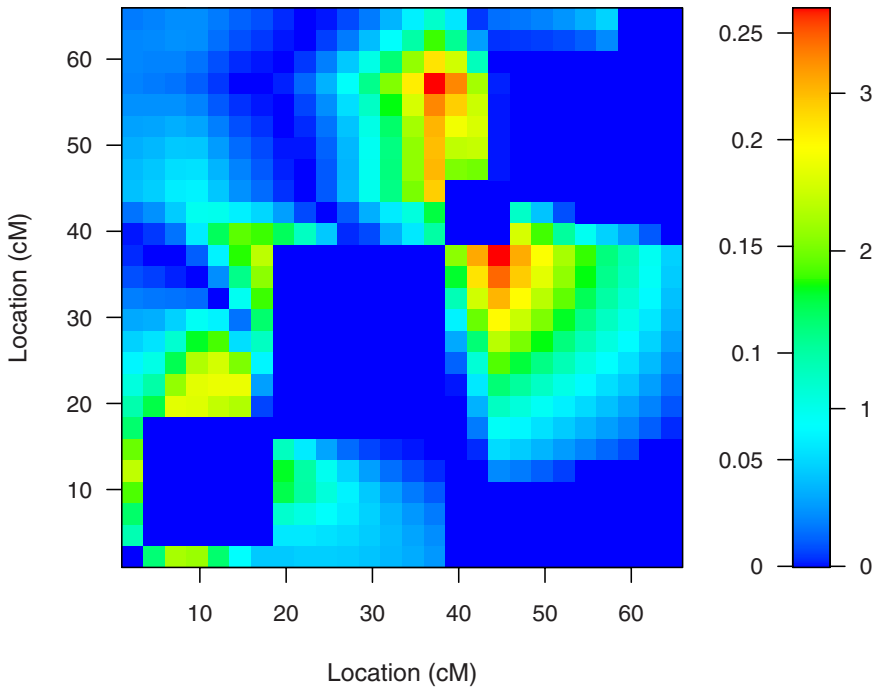


Figure 8.5. LOD scores, for chromosome 3, from a two-dimensional, two-QTL genome scan with the `hyper` data, with values for pairs of positions that are not separated by a marker replaced by 0. LOD_i is displayed in the upper left triangle; LOD_{av1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_{av1} , respectively.

We now use `find.marker` to find the names of the markers nearest the two inferred QTL, for use in `plot.pwg`. For the function `effectplot`, we may consider “pseudomarkers” (that is, the positions on the grid between markers), which we refer to directly by their chromosome and cM position. For example, we use “3@37.2” to refer to the pseudomarker closest to 37.2 cM on chromosome 3.

```
> mar <- find.marker(hyperc3, "3", c(37.2, 44.7))
```

Now we make the plots, which appear in Fig. 8.6. Note the use of `ylim` to adjust the *y*-axis limits in `effectplot`, so that the limits in the two plots correspond.

```
> par(mfrow=c(1,2))
> plot.pwg(hyperc3, marker=mar)
> effectplot(hyperc3, mname1="3@37.2", mname2="3@44.7",
+           ylim=range(pull.pheno(hyperc3,1)))
```

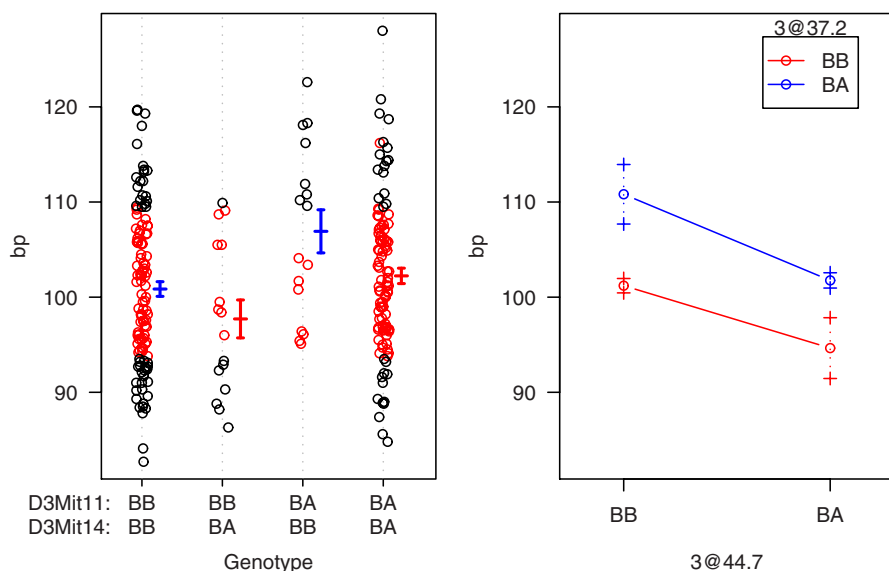


Figure 8.6. The effect of two putative linked QTL on chromosome 3 on the blood pressure phenotype in the **hyper** data. Left: a dot plot of the phenotype as a function of marker genotypes, with black dots corresponding to observed genotypes and red dots corresponding to missing (and so imputed) genotypes. Right: estimated phenotype averages for each of the four two-locus genotype groups, at the inferred locations of the two putative QTL.

Figure 8.6 may require a bit of study. In the left panel, with the output from `plot.pwg`, the red dots correspond to individuals missing genotypes at one or both markers (their two-locus genotypes are imputed from the available data), while the black dots correspond to individuals with genotype data at both markers. Note that the two recombinant classes have quite different phenotype averages: individuals that are BB at D3Mit11 and BA at D3Mit14 have low blood pressure, while those that are BA at D3Mit11 and BB at D3Mit14 have high blood pressure. The same feature is seen in the output from `effectplot`: the two QTL appear to have effects of opposite sign, but of approximately the same magnitude. Two linked QTL that have effects of opposite sign are said to be linked in *repulsion*. With QTL linked in repulsion, the region will exhibit little marginal effect, but if both loci are considered, the two may stand out clearly. With QTL linked in *coupling* (having effects of the same sign), on the other hand, a marginal effect will be apparent, but it will be difficult to separate the two loci.

We should, of course, also study the effects of the loci on chromosomes 1, 4, 6, and 15. We will again use `sim.geno` to impute the missing genotype data, and `effectplot` to plot the estimated phenotype averages as a function of QTL genotypes, considering just two QTL at a time.

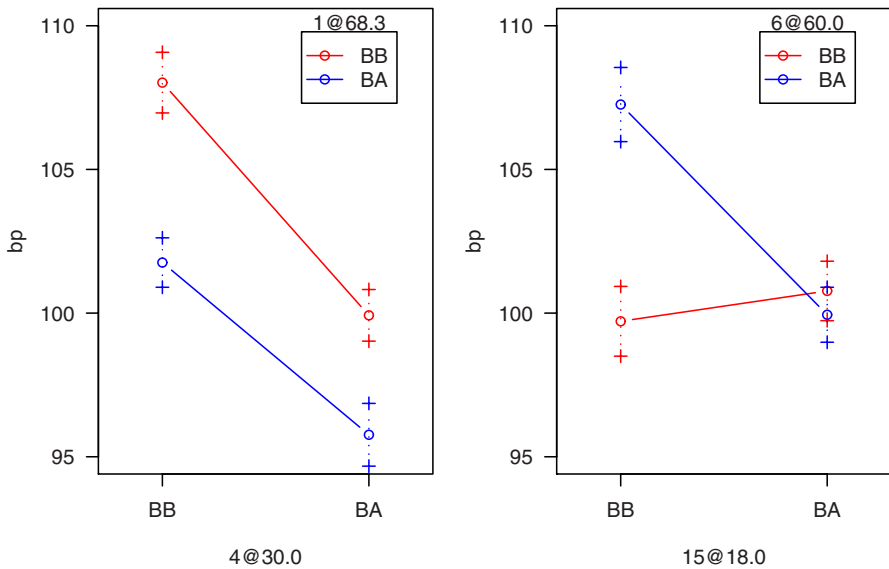


Figure 8.7. Estimated average blood pressure as a function of genotype at loci on chromosomes 1 and 4 (left panel) or chromosomes 6 and 15 (right panel), for the `hyper` data.

```
> hypersub <- sim.geno(subset(hyper, chr=c(1,4,6,15)), step=2.5,
+                      error.prob=0.001, n.draws=256)
> par(mfrow=c(1,2))
> effectplot(hypersub, mname1="1@68.3", mname2="4@30",
+            ylim=c(95, 110))
> effectplot(hypersub, mname1="6@60", mname2="15@18",
+            ylim=c(95, 110))
```

As seen in Fig. 8.7, the QTL on chromosomes 1 and 4 are seen to act approximately additively: the effect of the chromosome 4 locus is the same for each of the two genotypes at the chromosome 1 locus, and vice versa. Also, both loci have effects of the same sign: BB individuals have higher blood pressure than BA individuals.

The chromosome 6 and 15 loci, on the other hand, show clear epistasis: the chromosome 15 locus has effect only in the presence of the BA genotype at chromosome 6. Similarly, the chromosome 6 locus has effect only in the presence of the BB genotype at chromosome 15. In fact, only individuals that are BB at chromosome 15 and BA at chromosome 6 show high blood pressure; the other three genotype groups have similar, low blood pressure levels.

8.2 Binary traits

While the discussion in the previous section assumed a continuously varying phenotype, with residual variation following a normal distribution, the same techniques may be applied in the case of a binary phenotype.

In the full model (with two interacting QTL), we allow a separate penetrance for each two-locus genotype. That is, $\Pr(y = 1|q_1, q_2)$ is estimated separately for each possible pair of QTL genotypes (q_1, q_2) . If complete genotype data were available at the two putative QTL, the penetrance would be estimated by the proportion of individuals with that particular two-locus genotype that were affected.

In the additive QTL model, on the other hand, one must introduce a link function, as for the use of covariates with a binary trait (Sec. 7.3). In R/qtl, the *logit* link function, $\ln[\pi/(1 - \pi)]$, is used. And so, taking $\pi = \Pr(y = 1|q_1, q_2)$, the additive model is then

$$\ln[\pi/(1 - \pi)] = \mu + \beta_1 q_1 + \beta_2 q_2 \quad (8.4)$$

The meaning of “additive” (and so the meaning of “interaction” or “epistasis”) depends on the particular link function, just as, for a continuously varying phenotype, the meaning depends on the choice of transformation. With the logit link, we assume, in the additive model, that the effect of a change in the genotype at the first QTL is to modify the log odds for being affected by a constant multiplicative factor, independent of the genotype at the second QTL. (The *odds* corresponding to a probability π is the ratio $\pi/(1 - \pi)$. And so with a probability $\pi = 2/3$, the odds are 2:1.)

Aside from this need for a link function, the two-dimensional, two-QTL genome scan for a binary phenotype is directly analogous to that for a continuously varying phenotype with the normality assumption. We calculate the same sorts of LOD scores, and we use the same techniques to interpret the results. The link function is used to accommodate the fact that the penetrance, $\Pr(y = 1|q_1, q_2)$, takes values between 0 and 1, while the right-hand side of equation (8.4) need not be between 0 and 1. The link function transforms the penetrance so that both sides of equation (8.4) are unbounded.

Example

We return to the **nf1** data of Reilly *et al.* (2006), concerning neurofibromatosis type 1 (see Sec. 7.3). This is a backcross, with all individuals carrying the *NPcis* mutation, which may have come from their mother or from their father.

We must load the R/qtlbook package, and then the data, and then we drop one marker with completely missing genotype data.

```
> library(qtlbook)
> data(nf1)
> nf1 <- drop.nullmarkers(nf1)
```

In the single-QTL scan results (see Fig. 7.13 on page 203), we saw a big difference between individuals receiving the *NPcis* mutation from their mother (which showed linkage to chromosome 15) and those receiving the mutation from their father (which showed linkage to chromosome 19). Thus, we will perform the two-dimensional scan separately in these groups.

We first calculate the conditional QTL genotype probabilities (we use a coarse grid, for the sake of computational speed), and pull out the indicator for the origin of the *NPcis* mutation from the phenotype data.

```
> nf1 <- calc.genoprob(nf1, step=2.5, error.prob=0.001)
> frommom <- pull.pheno(nf1,"frommom")
```

Now we run the two-dimensional, two-QTL genome scans. Note the use of `model="binary"` to indicate analysis for a binary trait, which must take values 0 and 1. Also note that R/qtl uses maximum likelihood via the EM algorithm; while Haley–Knott regression or multiple imputation could also be used to deal with missing genotype data for a binary trait, these approaches have not yet been implemented in R/qtl.

```
> out.frommom <- scantwo(subset(nf1, ind=(frommom==1)),
+                         model="binary")
> out.fromdad <- scantwo(subset(nf1, ind=(frommom==0)),
+                         model="binary")
```

Before proceeding further, let us perform permutation tests, so that we may make sense of the results. The computations take a very long time, and so one may wish to split the calculations across multiple computers (see Sec. 8.1, page 223).

```
> operm.frommom <- scantwo(subset(nf1, ind=frommom==1),
+                         model="binary", n.perm=1000)
> operm.fromdad <- scantwo(subset(nf1, ind=frommom==0),
+                         model="binary", n.perm=1000)
```

Significance thresholds may be obtained with the `summary.scantwoperm` function. Note that these are a bit smaller than those obtained in Sec. 8.1 (for the *hyper* data, with a normal model).

```
> summary(operm.frommom)

affected (1000 permutations)
      full  fv1  int  add  av1
5%   5.68  4.46  4.13  4.54  2.28
10%  5.39  4.11  3.82  4.09  2.06

> summary(operm.fromdad)

affected (1000 permutations)
      full  fv1  int  add  av1
5%   5.58  4.40  4.11  4.48  2.25
10%  5.26  4.09  3.74  4.08  2.02
```

Let us study the results for the subset receiving the *NPcis* mutation from their mother. We will use a significance level of $\alpha = 0.2$ as a cutoff, ignoring the interaction LOD score, M_i .

```
> summary(out.frommom, perms=operm.frommom, pvalues=TRUE,
+         alphas=c(0.2, 0.2, 0, 0.2, 0.2))
```

	pos1f	pos2f	lod.full	pval	lod.fv1	pval	lod.int	pval
c7:c17	45	3	5.74	0.046	4.25	0.076	3.62	0.151
	pos1a	pos2a	lod.add	pval	lod.av1	pval		
c7:c17	40	43	2.11	0.915	0.622	1		

We see some evidence for interacting QTL on chromosomes 7 and 17. Note that linkage was previously seen to chromosome 15; it does not show up in this summary, as no one locus, when added to the chromosome 15 locus, sufficiently improves the fit. The chromosome 7 and 17 loci appear interesting only when considered together.

We may plot the two-dimensional scan results for chromosomes 7, 15 and 17 as follows. The result is in Fig. 8.8. LOD_i appears in the upper left triangle, and LOD_{fv1} appears in the lower right.

```
> plot(out.frommom, chr=c(7,15,17), lower="cond-int")
```

Let us now turn to the results for the subset receiving the *NPcis* mutation from their father. We again use a significance level of $\alpha = 0.2$ as a cutoff.

```
> summary(out.fromdad, perms=operm.fromdad, pvalues=TRUE,
+         alphas=c(0.2, 0.2, 0, 0.2, 0.2))
```

	pos1f	pos2f	lod.full	pval	lod.fv1	pval	lod.int	pval
c9:c19	58	0	5.52	0.055	2.53	0.933	0.504	1
	pos1a	pos2a	lod.add	pval	lod.av1	pval		
c9:c19	55.5	0	5.02	0.018	2.03	0.095		

We see evidence for a QTL on chromosome 9, in addition to the QTL previously identified on chromosome 19. The loci show no evidence of interaction. We plot LOD_i and LOD_{av1} for this pair of chromosomes as follows. See Fig. 8.9. The evidence for an interaction is negligible.

```
> plot(out.fromdad, chr=c(9, 19), lower="cond-add")
```

Let us complete our investigations with plots of the effects of the putative QTL pairs. We will use `effectplot`, though it is not ideal for a binary outcome. We first use `sim.geno` to perform multiple imputations of the genotypes at the putative QTL.

```
> nf1.fm <- sim.geno(subset(nf1, chr=c(7,17), ind=(frommom==1)),
+                   step=2.5, error.prob=0.001, n.draws=256)
> nf1.fd <- sim.geno(subset(nf1, chr=c(9,19), ind=(frommom==0)),
+                   step=2.5, error.prob=0.001, n.draws=256)
```

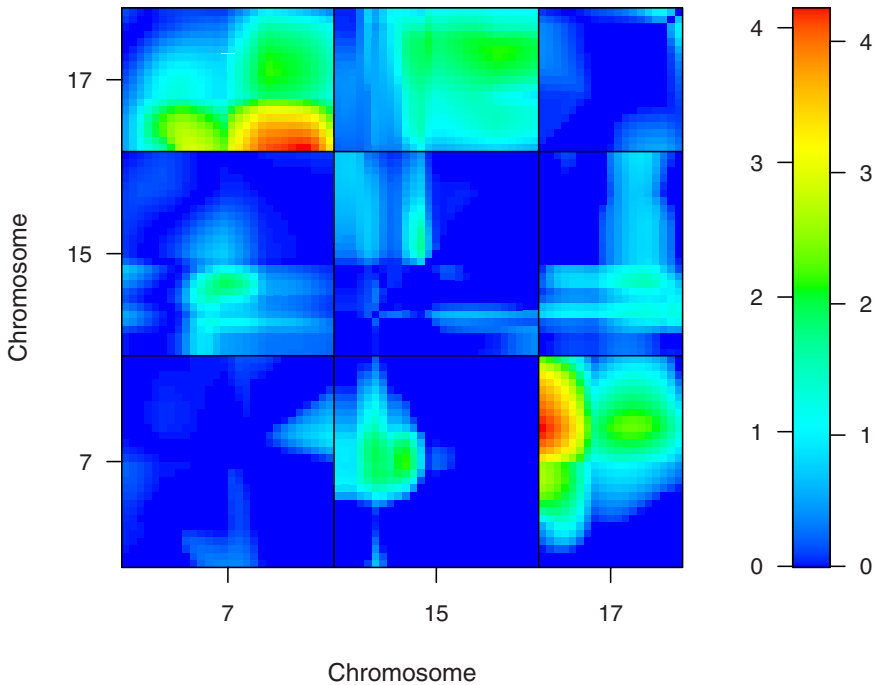


Figure 8.8. LOD scores, for selected chromosomes, from a two-dimensional, two-QTL genome scan with the *nf1* data, for those individuals receiving the *NPcis* mutation from their mother. LOD_i is displayed in the upper left triangle; LOD_{fv1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_{fv1} , respectively.

We now create the plots, which appear in Fig. 8.10. We refer to pseudo-markers (that is, the positions on the grid between markers), directly by their chromosome and cM position.

```
> par(mfrow=c(1,2))
> effectplot(nf1.fm, mname1="7@45", mname2="17@3",
+           ylim=c(0,1))
> effectplot(nf1.fd, mname1="9@55.5", mname2="19@0",
+           ylim=c(0,1))
```

In the left panel of Fig. 8.10, we see that, for the individuals receiving the *NPcis* mutation from their mother, those who are BB at both the chromosome 7 and 17 loci are largely unaffected, while in the other three groups, about half are affected. The right panel shows that, for the individuals receiving the mutation from their father, the putative QTL on chromosomes 9 and 19 display approximately additive effects. (However, recall that with the logit link, additivity is in terms of log odds and so would not give parallel lines in this plot.)

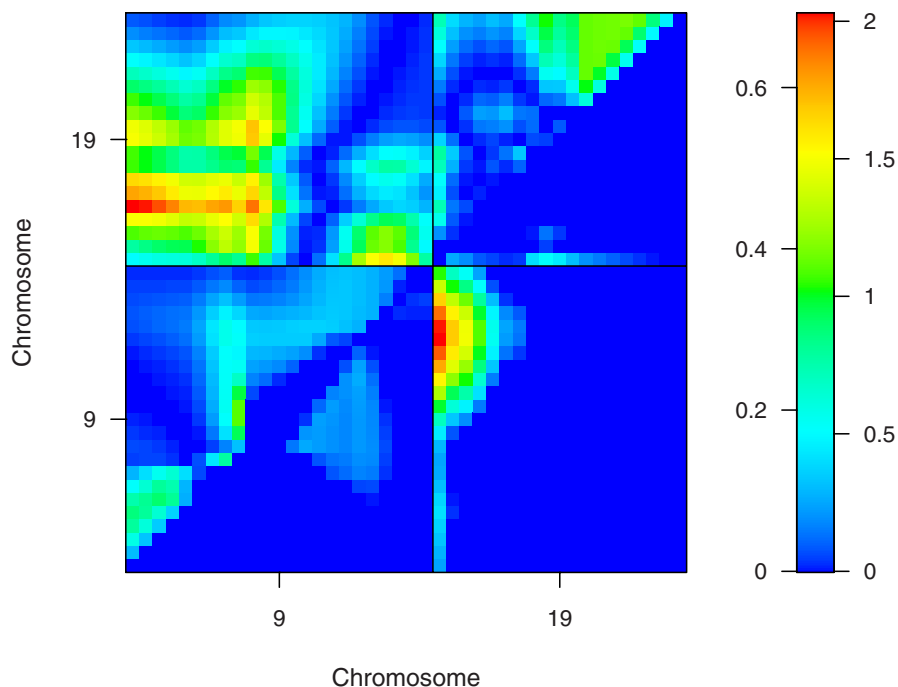


Figure 8.9. LOD scores, for selected chromosomes, from a two-dimensional, two-QTL genome scan with the *nf1* data, for those individuals receiving the *NPcis* mutation from their father. LOD_i is displayed in the upper left triangle; LOD_{av1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_{av1} , respectively.

8.3 The X chromosome

If you thought the discussion of the X chromosome in single-QTL analysis (Sec. 4.4) was painful, you may wish to skip this section. In a two-dimensional, two-QTL genome scan, the two-dimensional scan within the X chromosome needs to be treated specially, and also the scans for the X chromosome against each autosome require special treatment. There are several technical difficulties.

First, as in the case of the single-QTL scan, additional covariates may be needed under the null hypothesis (see Table 4.3 on page 112), in order to avoid spurious linkage to the X chromosome as a result of sex- or cross-direction-differences in the phenotype. These will be needed for the case of two QTL on the X chromosome, and for the case of one QTL on the X chromosome and one on an autosome. Moreover, in the latter situation, we require the results of the single-QTL scan on the autosomes with these covariates included, for the comparison of the two-QTL models to a single-QTL model.

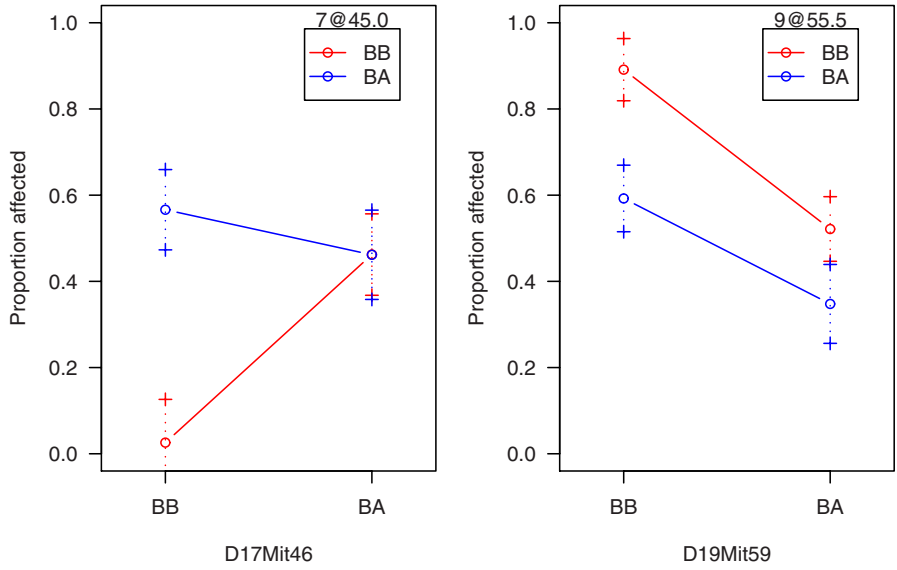


Figure 8.10. Estimated proportions of affected individuals as a function of genotype at two putative QTL for individuals receiving the *NPcis* mutation from their mother (left panel) or from their father (right panel), for the *nf1* data.

Table 8.1. Observed two-locus genotypes on the X chromosome, for a backcross with both sexes.

QTL 2	QTL 1			
	AA	AB	AY	BY
AA	*	*		
AB	*	*		
AY			*	*
BY			*	*

Second, in the case of two QTL on the X chromosome, we must recognize that not all two-locus genotypes will be observed. For example, consider the case of a backcross with both sexes. Females will have genotype AA or AB at a locus on the X chromosome, and males will be hemizygous A or B. In a single-QTL scan on the X, we consider these four groups, AA:AB:AY:BY. But in the consideration of two QTL on the X chromosome, only eight of the sixteen two-locus genotypes can actually be observed. (See Table 8.1.) While this is not a complex issue, the extension of algorithms for use with the X chromosome does require some care.

Finally, the degrees of freedom associated with our various LOD score statistics can vary greatly among the three possible cases: both putative QTL are on autosomes, one QTL is on an autosome and one is on the X chromosome, or both QTL are on the X chromosome. For example, consider our

Table 8.2. The number of estimated parameters for each model and the number of degrees of freedom for each test statistic, for a two-QTL scan in the case of an intercross with both sexes and both directions, according to the chromosome types for the two putative QTL.

chr type	No. parameters				No. df				
	f	a	1	0	f	fv1	i	a	av1
A:A	9	5	3	1	8	6	4	4	2
A:X	18	8	6	3	15	12	10	8	2
X:X	12	9	6	3	9	3	3	6	3

most complex situation, of an intercross with both sexes and both cross directions. The number of parameters for each of the four possible models and the number of degrees of freedom for each of the five possible test statistics, are displayed in Table 8.2. In the comparison of the full model (with two interacting QTL) to the best single-QTL model, with the M_{fv1} statistic, there are six degrees of freedom if both putative QTL reside on autosomes, three degrees of freedom if both putative QTL reside on the X chromosome, and twelve degrees of freedom if one QTL is on an autosome and one is on the X chromosome.

Thus, in assessing the statistical significance of the results of a two-dimensional, two-QTL scan, one must consider these three cases (illustrated in Fig. 8.11) separately, just as the autosomes and the X chromosome were considered separately in the single-QTL analysis (Sec. 4.4.2). We must admit that the details on how this should be done have not yet been worked out. For single-QTL analysis, we considered the autosomal and X chromosome genetic lengths. In the two-dimensional, two-QTL scan, one might consider the *areas* of the relevant regions.

Example

Let us briefly return to the `gutlength` data, described in Sec. 7.1. These data concern an intercross with both sexes and both directions.

We first load the data (which are contained in the R/qtlbook package), and run `calc.genoprob`. We will use a 5 cM grid, for the sake of speed, and because we will be taking only a cursory look at the results.

```
> data(gutlength)
> gutlength <- calc.genoprob(gutlength, step=5,
+                             error.prob=0.001)
```

We now run `scantwo`, using the defaults (maximum likelihood via the EM algorithm for the normal model).

```
> out.gl <- scantwo(gutlength)
```

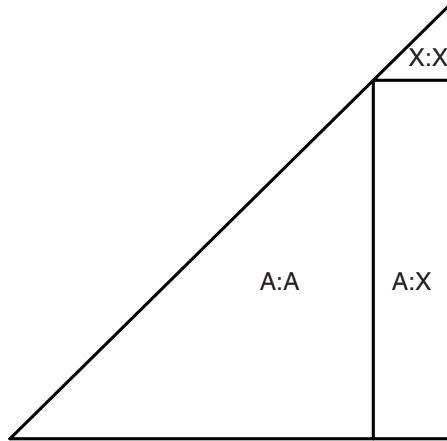


Figure 8.11. Regions in a two-dimensional, two-QTL scan, with the X chromosome included, that require separate treatment.

A plot of LOD_{fv1} and LOD_i is obtained as follows. (See Fig. 8.12.) We use `alternate.chrid=TRUE` so that the chromosome IDs may be more easily distinguished.

```
> plot(out.gl, lower="cond-int", alternate.chrid=TRUE)
```

Note the high LOD scores for the X chromosome when considered with any other chromosome: the segments for the X chromosome, on the top and right edges of the plot, really stand out. This is largely due to the fact that the degrees of freedom for these LOD score statistics are much larger for the A:X case (see Table 8.2). These portions of the two-dimensional scan require separate significance thresholds.

If we apply the simulation-derived significance thresholds for an intercross, cited in Sec. 8.1, we obtain the following summary.

```
> summary(out.gl, thresholds=c(9.1, 7.1, 6.3, 6.3, 3.3))
```

	pos1f	pos2f	lod.full	lod.fv1	lod.int	pos1a	pos2a
c5:cX	20	55	11.62	5.19	1.671	20	60
cX:cX	70	75	7.82	4.51	0.839	70	75
	lod.add	lod.av1					
c5:cX	9.95	3.52					
cX:cX	6.98	3.67					

We see evidence for QTL on chromosomes 5 and X (seen previously in the single-QTL scan; see Fig. 7.5 on page 188), and possibly two linked QTL on the X chromosome. But the significance thresholds we have used are likely inappropriate, and so these results should be viewed with a great deal of skepticism.

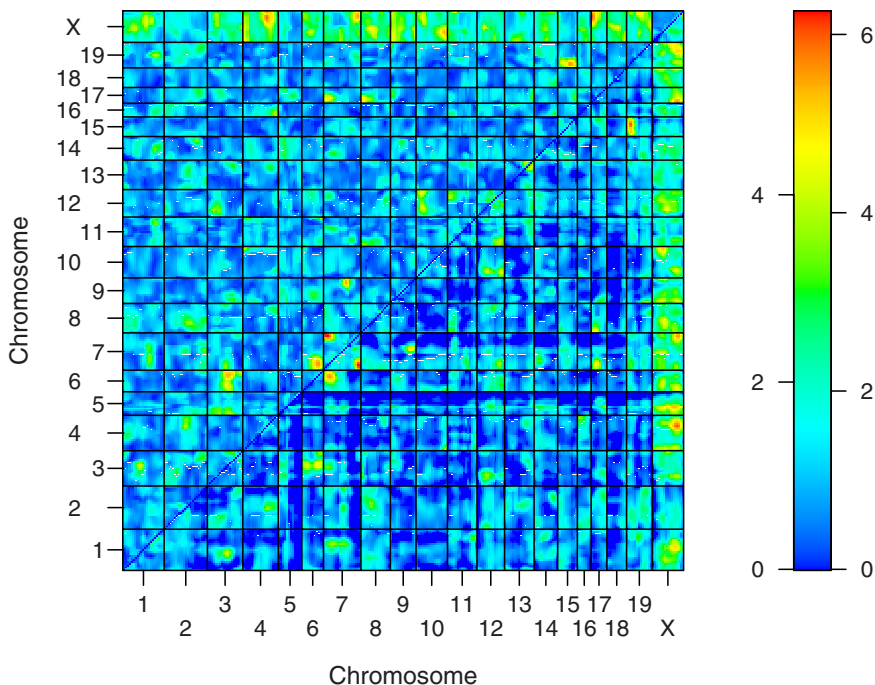


Figure 8.12. LOD scores from a two-dimensional, two-QTL genome scan with the *gutlength* data. LOD_i is displayed in the upper left triangle; LOD_{fv1} is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_{fv1} , respectively.

8.4 Covariates

As discussed in Chap. 7, it may be useful to take account of covariates, such as sex, in QTL analyses. If a covariate is associated with the phenotype, its consideration will reduce residual variation and so can enhance our ability to detect QTL. Additive covariates may be easily incorporated into a two-dimensional, two-QTL genome scan. With inclusion of an additive covariate, x , the four models in equation (8.1) (page 214) become the following.

$$\begin{aligned}
 H_f: y &= \mu + \beta_x x + \beta_1 q_1 + \beta_2 q_2 + \gamma(q_1 \times q_2) + \epsilon \\
 H_a: y &= \mu + \beta_x x + \beta_1 q_1 + \beta_2 q_2 + \epsilon \\
 H_1: y &= \mu + \beta_x x + \beta_1 q_1 + \epsilon \\
 H_0: y &= \mu + \beta_x x + \epsilon
 \end{aligned} \tag{8.5}$$

LOD scores may be calculated as before (see Sec. 8.1), and the interpretation of the two-dimensional scan results are essentially unchanged. (As discussed in Sec. 7.1, one should be cautious about the use of secondary phenotypes as covariates, as they will not necessarily be independent of QTL genotype.)

On the other hand, interactive covariates in a two-dimensional scan can be cumbersome, and the results are not easy to interpret. In R/qtl, interactive covariates in `scantwo`, indicated via the `intcovar` argument, are allowed to interact with both QTL, as well as with the QTL $\times$ QTL interaction. And so the four models become the following.

$$\begin{aligned}
 H_f: y &= \mu + \beta_x x + \beta_1 q_1 + \beta_2 q_2 + \gamma_{12}(q_1 \times q_2) + \\
 &\quad \gamma_{x1}(x \times q_1) + \gamma_{x2}(x \times q_2) + \gamma_{x12}(x \times q_1 \times q_2) + \epsilon \\
 H_a: y &= \mu + \beta_x x + \beta_1 q_1 + \beta_2 q_2 + \gamma_{x1}(x \times q_1) + \gamma_{x2}(x \times q_2) + \epsilon \\
 H_1: y &= \mu + \beta_x x + \beta_1 q_1 + \gamma_{x1}(x \times q_1) + \epsilon \\
 H_0: y &= \mu + \beta_x x + \epsilon
 \end{aligned} \tag{8.6}$$

The five LOD scores may be calculated as before, but the interpretation is rather different, as the LOD scores incorporate evidence for QTL $\times$ covariate interactions. Consider, for example, a binary covariate, such as sex, coded as $x = 0$ (female) or 1 (male). The interaction LOD score, LOD_i , being large indicates evidence for a QTL $\times$ QTL interaction in at least one of the two sexes.

We seldom make use of interactive covariates in a two-dimensional, two-QTL genome scan. While they may be useful, we prefer to postpone further investigation of QTL $\times$ covariate interactions to the more general exploration of multiple-QTL models, to be discussed in the next chapter.

Example

We return to the `gutlength` data, considered in the previous section and described in Sec. 7.1. First, we construct the covariates that we had used in Chap. 7.

```

> cross <- as.numeric(pull.pheno(gutlength, "cross"))
> frommom <- as.numeric(cross < 3)
> forw <- as.numeric(cross == 1 | cross == 3)
> sex <- as.numeric(pull.pheno(gutlength, "sex") == "M")
> crossX <- cbind(frommom, forw, frommom*forw)
> x <- cbind(sex, crossX)

```

Now, we perform the two-dimensional, two-QTL genome scan, with these additive covariates. It may take quite a bit of time.

```

> out.gl.a <- scantwo(gutlength, addcovar=x)

```

Let us plot the differences in the LOD scores obtained with and without the use of the additive covariates. We use `allow.neg=TRUE` so that the range of values includes negative numbers. Note that the differences are calculated via the function `-.scantwo`.

```

> plot(out.gl.a - out.gl, allow.neg=TRUE, alternate.chrid=TRUE)

```

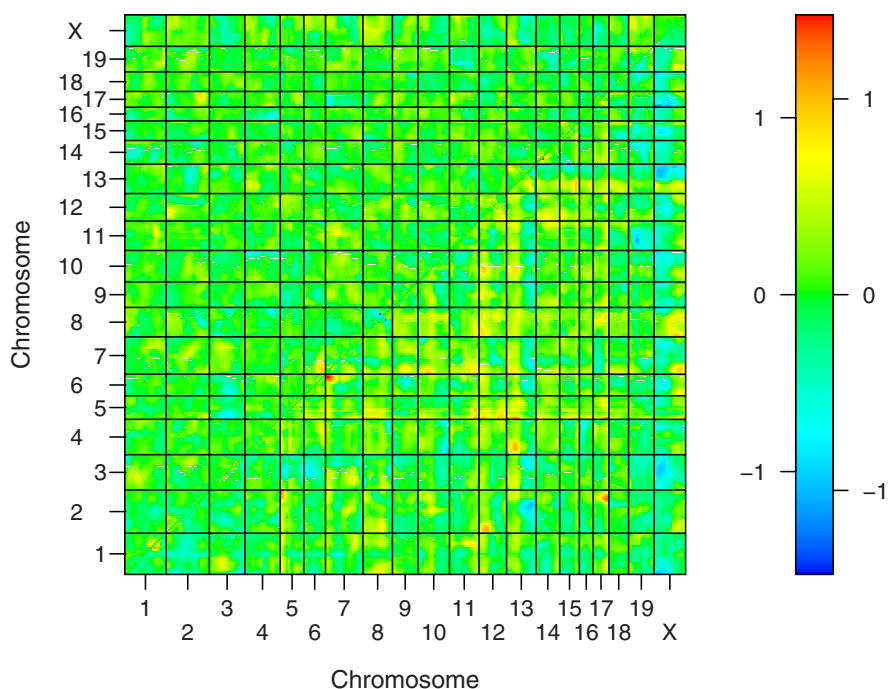


Figure 8.13. Differences in LOD scores calculated with and without the use of covariates, from a two-dimensional, two-QTL genome scan with the `gutlength` data. Differences in LOD_i are displayed in the upper left triangle; and differences in LOD_f are displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD_i and LOD_f , respectively.

Note that the LOD scores have gone up and down by as much as one unit (Fig. 8.13). However, the location of the maximum LOD score is the same, and has not changed much with the inclusion of the covariates. (We use the function `max.scantwo`, for pulling out the peak with maximum LOD score from a two-dimensional, two-QTL genome scan.)

```
> max(out.gl)
```

```
      pos1f pos2f lod.full lod.fv1 lod.int      pos1a pos2a
c5:cX    20   55   11.6    5.19    1.67        20    60
      lod.add lod.av1
c5:cX     9.95    3.52
```

```
> max(out.gl.a)
```

```
      pos1f pos2f lod.full lod.fv1 lod.int      pos1a pos2a
c5:cX    20   55   11.6    5.06    1.44        20    60
      lod.add lod.av1
c5:cX    10.2    3.61
```

8.5 Summary

Most QTL will be seen in the results of the traditional, single-QTL genome scan. However, two-dimensional, two-QTL genome scans, while computationally intensive, provide the opportunity to identify additional QTL, particularly those involved in epistatic interactions. In addition, the two-dimensional scan can give more clear evidence regarding linked QTL. This is particularly true in the case that two QTL are linked in *repulsion* (having effects of opposite sign): such QTL will show little marginal effect, but may stand out clearly when the two loci are considered jointly.

The interpretation of the results of a two-dimensional scan is not simple. We have described one approach to summarize the results, with an emphasis on measuring evidence for the presence of a second QTL. It is important to emphasize that the results of a two-dimensional scan need to be considered in combination with the initial single-QTL analysis. Moreover, the findings should be considered preliminary, to be used as a starting point for the fit and exploration of multiple-QTL models.

The X chromosome presents additional difficulties in the context of a two-dimensional genome scan. The varying degrees of freedom attached to different LOD scores (in the case of two QTL on autosomes, two QTL on the X chromosome, or one QTL on an autosome and one on the X chromosome) imply that separate significance thresholds will be required, but how this should be done remains a matter for research.

It is often important to consider covariates within QTL analyses. The inclusion of additive covariates in a two-dimensional, two-QTL genome scan is simple. Interactive covariates, however, are difficult to work with in this context, and so we have seldom made use of them in two-dimensional scans.

8.6 Further reading

Haley and Knott (1992) were the first to propose an exhaustive search over all two-QTL models. Sugiyama *et al.* (2001) applied the approach to the **hyper** data, which we have used extensively as an example. Sen and Churchill (2001) helped to solidify the use of two-dimensional, two-QTL scans as a general technique. Ljungberg *et al.* (2004) described an algorithm for identifying the optimal pair of loci in a two-dimensional scan without performing an exhaustive search.