Chapter 8

Markov chain Monte Carlo on Pedigrees

8.1 Simulation conditional on data: MCMC

Equation (7.10) gave the likelihood for a genetic model on a pedigree as an expectation over latent variables $\mathbf{X}$, and hence, in principle, provided a method for Monte Carlo estimation of the likelihood. We need to estimate

$$L(\theta) \;=\; P_\theta(\mathbf{Y}) \;=\; \sum_{\mathbf{X}} P_\theta(\mathbf{X}, \mathbf{Y}).$$

As previously, any suitable latent variables may be used, normally either meiosis indicators $\mathbf{S}$ or genotypes $\mathbf{G}$. For convenience, we use the general notation $\mathbf{X}$ for the general formulation.

However, unless the simulation distribution $P^*(\mathbf{X})$ is conditioned in some way on data $\mathbf{Y}$, equation (7.10) is often useless. Genotypes or gene descent patterns simulated from the prior probability distribution given only the model and the pedigree structure will rarely even be consistent with the observed data. Importance sampling considerations dictate that the sampling distribution should be close to proportional to $P_\theta(\mathbf{X}, \mathbf{Y})$, or as a function of latent variables $\mathbf{X}$ to $P_\theta(\mathbf{X} \mid \mathbf{Y})$ (equation (7.12)). Intuitively also, to obtain realizations that have better than infinitesimal probability of giving a non-negligible contribution to the likelihood we must simulate conditional on the data. However

$$P_\theta(\mathbf{X} \mid \mathbf{Y}) \;=\; \frac{P_\theta(\mathbf{X}, \mathbf{Y})}{P_\theta(\mathbf{Y})}, \tag{8.1}$$

and the normalizing factor $P_\theta(\mathbf{Y})$ is unknown. If we could compute $L(\theta) = P_\theta(\mathbf{Y})$, Monte Carlo estimation of likelihoods would be unnecessary.

Enter Markov chain Monte Carlo, or MCMC. We review briefly the Metropolis-Hastings class of algorithms (Hastings, 1970) for generating dependent realizations from a target probability distribution known only up to a normalizing factor. For

consistency of notation, we denote the target distribution by $P_\theta(\mathbf{X} \mid \mathbf{Y})$. The space of possible values of $\mathbf{X}$ is denoted $\mathcal{X}$. For each $\mathbf{X}$ in $\mathcal{X}$ a *proposal distribution* $q(\cdot; \mathbf{X})$ is defined. Then, if the process is now at $\mathbf{X}$ the next value is generated as follows:

1. Generate $\mathbf{X}^\dagger$ from the proposal distribution $q(\cdot; \mathbf{X})$
2. Compute the Hastings ratio

$$h(\mathbf{X}^\dagger; \mathbf{X}) = \frac{q(\mathbf{X}; \mathbf{X}^\dagger) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})}{q(\mathbf{X}^\dagger; \mathbf{X}) P_\theta(\mathbf{X} \mid \mathbf{Y})}. \tag{8.2}$$

Note that h depends only on the ratio of densities $P_\theta(\cdot \mid \mathbf{Y})$, so that any normalizing factor need not be computed.

3. The resampled X^* is then determined from the Hastings ratio as follows:

$$\begin{aligned} P^*(\mathbf{X}^* = \mathbf{X}^\dagger) &= a = \min(1, h(\mathbf{X}^\dagger; \mathbf{X})) \\ P^*(\mathbf{X}^* = \mathbf{X}) &= (1 - a). \end{aligned}$$

Thus a is the *acceptance probability* for the proposed $\mathbf{X}^\dagger$.
Clearly, given the current value of $\mathbf{X}$, the probability distribution of $\mathbf{X}^*$ is determined, independently of the past of the process: a Markov chain on the space $\mathcal{X}$ of values of $\mathbf{X}$ has been defined.

It remains to show that the desired distribution $P_\theta(\mathbf{X}|\mathbf{Y})$ is an equilibrium distribution of the Markov chain. Hence, if the chain is aperiodic and irreducible, $P_\theta(\mathbf{X}|\mathbf{Y})$ is the unique equilibrium distribution. In this case, the ergodic theorem provides that time averages over realizations of the chain converge to expectations under the equilibrium distribution. These time-averages may then be used as Monte Carlo estimates of these expectations, just as previously in sections 3.7 and 7.6 simple averages of independent realizations were used.

The net resampling distribution $P^*(\mathbf{X}^*)$ is compounded from the proposal $q(\mathbf{X}^\dagger; \mathbf{X})$ and the acceptance or rejection step. Since the process is symmetric in $\mathbf{X}$ and a proposed $\mathbf{X}^\dagger$, with $h(\mathbf{X}^\dagger; \mathbf{X}) = (h(\mathbf{X}; \mathbf{X}^\dagger))^{-1}$, without loss of generality we can assume $h(\mathbf{X}^\dagger; \mathbf{X}) \geq 1$ or

$$q(\mathbf{X}; \mathbf{X}^\dagger) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y}) \geq q(\mathbf{X}^\dagger; \mathbf{X}) P_\theta(\mathbf{X} \mid \mathbf{Y}).$$

Then a proposed transition from $\mathbf{X}$ to $\mathbf{X}^\dagger$ is accepted ($a = 1$) and the probability of the move is the proposal probability:

$$P^*(\mathbf{X}^\dagger; \mathbf{X}) = q(\mathbf{X}^\dagger; \mathbf{X}).$$

For the reverse move, from $\mathbf{X}^\dagger$, $\mathbf{X}$ must be both proposed and accepted. Thus, the probability, $P^*(\mathbf{X}; \mathbf{X}^\dagger)$, of the reverse transition is

$$\begin{aligned} q(\mathbf{X}; \mathbf{X}^\dagger) h(\mathbf{X}; \mathbf{X}^\dagger) &= q(\mathbf{X}; \mathbf{X}^\dagger) \frac{q(\mathbf{X}^\dagger; \mathbf{X}) P_\theta(\mathbf{X} \mid \mathbf{Y})}{q(\mathbf{X}; \mathbf{X}^\dagger) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})} \\ &= q(\mathbf{X}^\dagger; \mathbf{X}) \frac{P_\theta(\mathbf{X} \mid \mathbf{Y})}{P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})}. \end{aligned}$$

Combining these two equations, we have

$$P^*(\mathbf{X}^\dagger; \mathbf{X}) P_\theta(\mathbf{X} \mid \mathbf{Y}) = P^*(\mathbf{X}; \mathbf{X}^\dagger) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y}). \tag{8.3}$$

In words, under the defined Markov chain and distribution $P_\theta(\cdot \mid \mathbf{Y})$, the probability of being at $\mathbf{X}$ and moving to $\mathbf{X}^\dagger$ is the same as the probability of being at $\mathbf{X}^\dagger$ and moving to $\mathbf{X}$. This *detailed balance* condition holds for all $\mathbf{X}$ and $\mathbf{X}^\dagger$, which is a sufficient condition for $P_\theta(\cdot \mid \mathbf{Y})$ to be an equilibrium distribution of the Markov chain.

The algorithm of Metropolis et al. (1953) is a special case; if $q(\mathbf{X}^\dagger; \mathbf{X}) = q(\mathbf{X}; \mathbf{X}^\dagger)$ the Hastings ratio reduces to the odds ratio of the proposal state $\mathbf{X}^\dagger$ versus the current state $\mathbf{X}$. An alternative version of MCMC sampling is the Gibbs sampler (Geman and Geman, 1984). We consider here the general case in which, at a given step, $\mathbf{X}$ is partitioned into two sets of components, $\mathbf{X} = (\mathbf{X}_u, \mathbf{X}_f)$, the subscripts u denoting *updated* and f denoting *fixed*. These subsets change at each step, so that every component of $\mathbf{X}$ is sometimes updated. The sampled $\mathbf{X}^*$ differs from $\mathbf{X}$ only in the set of components $\mathbf{X}_u$, and $\mathbf{X}_u^*$ is sampled from the distribution $P_\theta(\mathbf{X}_u|\mathbf{X}_f, \mathbf{Y})$. Suppose $\mathbf{X}$ is currently from the desired distribution $P_\theta(\mathbf{X} \mid \mathbf{Y})$, so that the marginal distribution of the current $\mathbf{X}_f$ is $P_\theta(\mathbf{X}_f \mid \mathbf{Y})$. Thus the distribution of the resampled $\mathbf{X}^*$ is

$$\begin{aligned} P^*(\mathbf{X}_u^*, \mathbf{X}_f^*) &= P^*(\mathbf{X}_u^* \mid \mathbf{X}_f^*) P^*(\mathbf{X}_f^*) \\ &= P_\theta(\mathbf{X}_u^*|\mathbf{X}_f, \mathbf{Y}) P_\theta(\mathbf{X}_f \mid \mathbf{Y}) \\ &= P_\theta(\mathbf{X}^* \mid \mathbf{Y}). \end{aligned} \tag{8.4}$$

Thus the Gibbs sampler also maintains the equilibrium distribution $P_\theta(\cdot \mid \mathbf{Y})$.

The Gibbs sampler is, in fact, a special case of a Metropolis-Hastings sampler. Consider a Metropolis-Hastings sampler in which the proposal distribution is the resampling distribution of the Gibbs sampler:

$$q(\mathbf{X}^\dagger; \mathbf{X}) = P_\theta(\mathbf{X}_u^\dagger|\mathbf{X}_f, \mathbf{Y}) I(\mathbf{X}_f^\dagger \equiv \mathbf{X}_f)$$

where $I(\cdot)$ is the indicator function. Then the Hastings ratio is

$$\begin{aligned} h(\mathbf{X}^\dagger; \mathbf{X}) &= \frac{q(\mathbf{X}; \mathbf{X}^\dagger) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})}{q(\mathbf{X}^\dagger; \mathbf{X}) P_\theta(\mathbf{X} \mid \mathbf{Y})} \\ &= \frac{P_\theta(\mathbf{X}_u|\mathbf{X}_f, \mathbf{Y}) P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})}{P_\theta(\mathbf{X}_u^\dagger|\mathbf{X}_f, \mathbf{Y}) P_\theta(\mathbf{X} \mid \mathbf{Y})} \\ &= \frac{P_\theta(\mathbf{X} \mid \mathbf{Y})}{P_\theta(\mathbf{X}_f \mid \mathbf{Y})} \frac{P_\theta(\mathbf{X}_f \mid \mathbf{Y})}{P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})} \frac{P_\theta(\mathbf{X}^\dagger \mid \mathbf{Y})}{P_\theta(\mathbf{X} \mid \mathbf{Y})} \\ &= 1. \end{aligned}$$

In this case $a \equiv h(\mathbf{X}^\dagger; \mathbf{X}) \equiv 1$, and no rejection step is necessary. Although, in the Gibbs sampler there is no rejection step, $\mathbf{X}^* = \mathbf{X}$ is possible, since $\mathbf{X}_u$ is a possible value for the resampled $\mathbf{X}_u^*$.

In order for the time-average over the chain to converge to the expectation under the equilibrium distribution, the ergodic theorem must apply. For discrete Markov chains, we need irreducibility. However, in practice, too much attention is paid to irreducibility. Any chain can be made irreducible, using Metropolis rejection, but irreducibility *per se* is useless. For example, one might decide that once in

a million trials one will propose a new realization from the prior distribution of latent variables. Once in a million million realizations one might get something compatible with the data. Once in a million, million, million trials one might get an accepted realization. Obviously nothing has changed with regard to realizations from the chain, but the sampler is irreducible. Metropolis rejected restarts are often a good idea — one of several key ideas in getting better samplers, and in assessing how good they are. However, it has to be done with the practical goal of more efficient Monte Carlo estimation.

There are two (related) sorts of convergence which often get confused. One is convergence of the marginal distribution of each $\mathbf{X}^{(\tau)}$ to the equilibrium distribution of the Markov chain as τ becomes large. The other relates to the convergence of a time-average over the chain to the expectation of the function under the equilibrium distribution. Both depend on the mixing properties of the Markov chain, and parameters such as the largest non-unit eigenvalue of the transition matrix, but the first can (in principle) be addressed by burn-in (discarding enough realizations before starting to accumulate the time-average) and is not normally a practical problem. The second class of questions remain even if we could start in the equilibrium probability distribution. This is a much bigger problem; all parts of the space contributing substantially to the target probability distribution must be sampled. Although shorter runs in different parts of the space may be helpful in diagnosing a problem, Monte Carlo estimation must be done using a time-average of a single realization of the Markov chain process. Runs in different parts of the space cannot be combined, without knowledge of how to weight the realizations from the different starts. (See Geyer (1992) for more discussion.)

Estimation of the standard deviations of Monte Carlo estimates of expectations is essential. Several easily implemented estimators have been proposed, but assessment of the estimates is hard, in practice. Again, Geyer (1992) is a good reference. One of the simplest methods of estimating Monte Carlo variances is by using batch means (Hastings, 1970). One divides the realizations into sufficiently large batches so that the batch means are "almost independent", and relates the variance of independent batch means to the variance of the overall mean (the estimator of the expectation). The variance of independent batch means can be estimated from the empirical variance as in section 3.7. One can test for autocorrelation between the batch means. This is quite effective if the sampler is doing well, but can severely underestimate variance if the sampler is not getting around the space. However, other variance estimators have the same deficiency, and the empirical variance of the batch means is easily computed.

Variance estimation also relates to the choice of spacing in sampling realizations from an MCMC. The optimal spacing is the one that achieves minimum computational cost for given precision of the resulting estimator. This optimal spacing depends on the relative costs of generating the samples and of evaluating the contribution to the estimator at the realized values, but is seldom large (Geyer, 1992).

This section has aimed only to outline the main principles and issues in MCMC. For those who wish to pursue the topic, Gilks et al. (1996) is a good starting point, while there is already a large more recent literature.

8.2 Single-site updating methods

As in other areas of application, the earliest MCMC samplers that were used to realize latent variables on pedigrees conditional on phenotypic data were mainly single-site updating methods. The proposed changes to the latent variable configurations were thus very small. Lange and Matthysse (1989) used as their latent variables both the genotypes and inheritance patterns of genes, and used a Metropolis algorithm to propose changes. Sheehan (1990) and Thompson and Guo (1991) used a Gibbs sampling approach, using the genotypes as the latent variables, while Thompson (1994*a*; 1994*b*) used a Metropolis algorithm to update a single meiosis indicator $S_{i,j}$ for meiosis i and locus j.

Unfortunately, in genetic examples the constraints on genotypes $\mathbf{G}$ or meiosis indicators $\mathbf{S}$ imposed by Mendelian segregation and discrete marker phenotypes mean that any proposal that makes multiple changes to the current value of $\mathbf{G}$ or $\mathbf{S}$ has a high probability of proposing a configuration inconsistent with the data $\mathbf{Y}$. By contrast, although proposed changes are small, single-site updates are easily proposed and often accepted. The genes and heritable effects in an individual are determined by those in his parents, and jointly with those in his spouse, influence those in his offspring (Figure 1.3(a)). This neighborhood structure means that a single-site Gibbs sampler is easy to implement. Each genetic effect in each individual is successively updated, conditional upon the remainder.

Specifically, where genotypes $\mathbf{G}$ are the latent variables, underlying genotypes for both trait and marker loci are sampled individual by individual and locus by locus. For a single-site update to component $G_{i,j}$, the genotype of individual i at locus j, the proposal distribution for the Gibbs sampler (equation (8.4)) is

$$\begin{aligned} q_{i,j}(\mathbf{G}^*;\mathbf{G}) &= P_\theta(G^*_{i,j} \mid \mathbf{G}_{-(i,j)}, \mathbf{Y}) \text{ for component } (i,j) \\ G^*_{k,l} &= G_{k,l} \text{ for } (k,l) \neq (i,j), \quad \text{or } \mathbf{G}^*_{-(i,j)} = \mathbf{G}_{-(i,j)}. \end{aligned} \tag{8.5}$$

As for $\mathbf{S}$ in section 4.7, we use the standard notation $\mathbf{G}_{-(i,j)}$ for the set of all components of $\mathbf{G}$ other than $G_{i,j}$. This full conditional distribution for $G_{i,j}$ is easily computed, but only small changes to $\mathbf{G}$ are possible at each step. On the other hand, the full conditionals for larger blocks of components $\mathbf{G}_\mathcal{T} = \{G_{i,j}; (i,j) \in \mathcal{T}\}$ are more computationally intensive or even infeasible.

For certain data configurations, the single-site genotypic Gibbs sampler is not irreducible when a locus is multiallelic. However, theoretical irreducibility can always be easily achieved. The practical problem is failure of the sampler to mix adequately. This can be a problem on large pedigrees even for diallelic loci, particularly if underlying genotypes are highly constrained (but not determined) by the data. The reducibility of the Gibbs sampler for genetic loci with more than two alleles was first addressed by Sheehan and Thomas (1993), in the context of a single-genotype Gibbs sampler. Their method used modification of either the segregation probabilities or the penetrance probabilities, so that the sampler was no longer irreducible. For example, modifying the penetrances

$$\begin{aligned} P^*(Y_{i,j} \mid G_{i,j}) &= P_\theta(Y_{i,j} \mid G_{i,j}) \quad \text{if } P_\theta(Y_{i,j} \mid G_{i,j}) > 0 \\ P^*(Y_{i,j} \mid G_{i,j}) &= c \qquad\qquad \text{if } P_\theta(Y_{i,j} \mid G_{i,j}) = 0. \end{aligned} \tag{8.6}$$

Then

$$\begin{aligned}\frac{P_\theta(\mathbf{G},\mathbf{Y})}{P^*(\mathbf{G},\mathbf{Y})} &= 1 \text{ if } P_\theta(\mathbf{Y} \mid \mathbf{G}) > 0 \\ &= 0 \text{ if } P_\theta(\mathbf{Y} \mid \mathbf{G}) = 0.\end{aligned}$$

Thus no reweighting is required in order for the realizations to represent the distribution of genotypes under the true genetic model. All realizations consistent with the true model have equal weight; those inconsistent with it are just dropped from the output sample. Lin et al. (1993) used similar penetrance modifications to achieve irreducibility, but used Metropolis-coupled samplers (Geyer, 1991*a*), coupling a sampler under the true model to samplers which were not only irreducible, but also moved more quickly around the space. Rather than a uniform penetrance modification for all individuals, only individual-specific changes necessary to achieve irreducibility are made. The expansion of the space that is sampled is therefore limited.

Several methods for more efficient sampling of the space of feasible underlying genotype configurations have been developed. Some of these are due to Shili Lin (Lin et al., 1993; Lin et al., 1994). Others are due to Eric Sobel (Sobel and Lange, 1993) and to Charles Geyer (Geyer and Thompson, 1995). We briefly outline here only the methods of Lin et al. (1993; 1994), directed specifically towards sampling of genotypes at polymorphic marker loci where there are many unsampled individuals in the pedigree. These methods use a form of "heated proposals", resulting in samplers that move around the space of genotypic configurations far more effectively.

One possibility is to base a Metropolis-Hastings sampler on the local conditional distribution for the single component $G_{i,j}$ (equation (8.5)), but in a way that enhances movement around the space. The method of Lin et al. (1994) "flattens" the proposal distribution in a manner similar to simulated annealing, using a "temperature" parameter T:

$$\begin{aligned}q_{i,j}(\mathbf{G}^*;\mathbf{G}) &\propto (P_\theta(G^*_{i,j} \mid \mathbf{G}_{-(i,j)},\mathbf{Y}))^{1/T} \text{ for component } (i,j) \\ G^*_{k,l} &= G_{k,l} \text{ for } (k,l) \neq (i,j), \quad \text{or } \mathbf{G}^*_{-(i,j)} = \mathbf{G}_{-(i,j)}.\end{aligned}$$

The Hastings ratio is then

$$\begin{aligned}h(\mathbf{G}^*;\mathbf{G}) &= \frac{q(\mathbf{G};\mathbf{G}^*)P_\theta(\mathbf{G}^* \mid \mathbf{Y})}{q(\mathbf{G}^*;\mathbf{G})P_\theta(\mathbf{G} \mid \mathbf{Y})} \\ &= \frac{(P_\theta(G_{i,j} \mid \mathbf{G}^*_{-(i,j)},\mathbf{Y}))^{1/T}P_\theta(\mathbf{G}^* \mid \mathbf{Y})}{(P_\theta(G^*_{i,j} \mid \mathbf{G}_{-(i,j)},\mathbf{Y}))^{1/T}P_\theta(\mathbf{G} \mid \mathbf{Y})} \\ &= \frac{(P_\theta(\mathbf{G} \mid \mathbf{Y}))^{1/T}P_\theta(\mathbf{G}^* \mid \mathbf{Y})(P_\theta(\mathbf{G}_{-(i,j)} \mid \mathbf{Y}))^{1/T}}{(P_\theta(\mathbf{G}^* \mid \mathbf{Y}))^{1/T}P_\theta(\mathbf{G} \mid \mathbf{Y})(P_\theta(\mathbf{G}^*_{-(i,j)} \mid \mathbf{Y}))^{1/T}} \\ &= \frac{(P_\theta(\mathbf{G}^* \mid \mathbf{Y}))^{1-1/T}}{(P_\theta(\mathbf{G} \mid \mathbf{Y}))^{1-1/T}} \\ &= \frac{(P_\theta(G^*_{i,j} \mid \mathbf{G}_{-(i,j)},\mathbf{Y}))^{1-1/T}}{(P_\theta(G_{i,j} \mid \mathbf{G}_{-(i,j)},\mathbf{Y}))^{1-1/T}}\end{aligned}$$

using, in several steps, the fact that $\mathbf{G}_{-(i,j)} = \mathbf{G}^*_{-(i,j)}$. The Hastings ratio is thus as easily computed as the local conditionals $P_\theta(G_{i,j} \mid \mathbf{G}_{-(i,j)}, \mathbf{Y})$. An interesting feature of this system is that, with $T > 1$, the probability of change in $\mathbf{G}$ is reduced from that for the Gibbs sampler, where $T = 1$ (C. Jennison, pers. comm. 1992). However, because this increases the probability that the sampler remains in low-probability states, it increases the overall probability of a succession of changes that moves $\mathbf{G}$ to a different part of the space. The probabilities of single-step changes are not necessarily indicative of overall performance of the sampler, particularly in high-dimensional spaces.

Under the assumption that $S_{\bullet,j}$ are first-order Markov over loci j (section 4.7), the single-site meiosis indicator sampler is also easily implemented (Thompson, 1994*a*). Since $S_{i,j}$ is binary, a Metropolis algorithm is natural. A meiosis i and locus j are selected at random, and a change from $S_{i,j} = s$ to $S_{i,j} = (1 - s)$ is proposed. This proposal changes only the recombinant/non-recombinant status in the two intervals adjoining locus j, and the conditional probability of marker data at locus j:

$$
\begin{aligned}
h(\mathbf{S}^*;\mathbf{S}) &= \frac{P_\theta(\mathbf{Y} \mid \mathbf{S}^*)P_\theta(\mathbf{S}^*)}{P_\theta(\mathbf{Y} \mid \mathbf{S})P_\theta(\mathbf{S})} \\
&= \frac{P_\theta(Y_{\bullet,j} \mid S^*_{\bullet,j})P_\theta(S^*_{i,j} \mid S_{i,j-1}, S_{i,j+1})}{P_\theta(Y_{\bullet,j} \mid S_{\bullet,j})P_\theta(S_{i,j} \mid S_{i,j-1}, S_{i,j+1})} \\
&= \frac{P_\theta(Y_{\bullet,j} \mid S^*_{\bullet,j})}{P_\theta(Y_{\bullet,j} \mid S_{\bullet,j})} \left(\frac{\rho_{j-1}}{1-\rho_{j-1}}\right)^{T_{j-1}} \left(\frac{\rho_j}{1-\rho_j}\right)^{T_j}, \qquad (8.7)
\end{aligned}
$$

for $j = 1, \ldots, L$ (see equation 4.12). Here $\rho_{j-1} = \Pr(S_{i,j-1} \neq S_{i,j})$ is the recombination frequency between locus $j-1$ and locus j, and $T_{j-1} = (|S_{i,j-1} - s| - |S_{i,j-1} - 1 + s|)$ is the indicator of whether the proposal places ($T_{j-1} = +1$) or removes ($T_{j-1} = -1$) a recombination between locus $j-1$ and j. The values ρ_j and T_j are analogously defined for the interval j to $j+1$, and $\rho_0 = \rho_L = \frac{1}{2}$. The first term in the Hastings ratio $h(\mathbf{S}^*;\mathbf{S})$ is given by equation (3.10) and is easily computed by the methods outlined in that section, provided there are not too many data $S_{\bullet,j}$ on the pedigree. Generally, the space of latent variables is smaller for $\mathbf{S}$ than for $\mathbf{G}$, and hence MCMC is more effective. The sampler may not be irreducible (Sobel and Lange, 1996), but there are many fewer constraints than with a genotypic sampler and irreducibility is often provable on a locus-by-locus basis (Thompson, 1994*a*; Thompson and Heath, 1999). Note that, provided recombination frequencies between adjacent loci are strictly positive, irreducibility is a single-locus issue.

8.3 Combining exact computation and Monte Carlo

A major difficulty with MCMC methods is to ensure proper mixing of the samplers, and hence efficient Monte Carlo estimation. On large pedigrees, with models

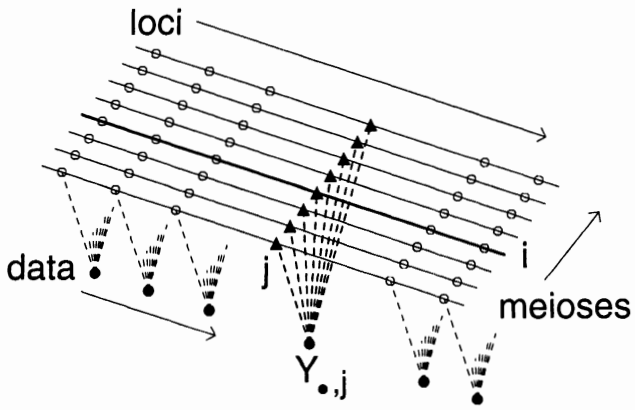

FIGURE 8.1. *The conditional independence structure for MCMC sampling*

or data involving multiple linked loci, single-variable MCMC updating methods are not effective. Some approaches to improving Monte Carlo estimates involve some combination of exact and Monte Carlo computation. One straightforward idea is simply to compute exactly on those parts of the pedigree on which this is possible (Thompson, 1991). The results from peeling peripheral parts of the pedigree enter as potentials on nodes of the remaining core (Geyer and Thompson, 1995), and the space over which MCMC sampling is required is reduced. Rao-Blackwellized estimators for mixed-model likelihoods (section 9.3) also combine exact computation with MCMC sampling. However, the sampling used for these estimators by Thompson and Guo (1991) was single-site updating. Major improvements can be gained only by improved MCMC samplers.

Recently a variety of joint-updating schemes have been developed. For example, Jensen et al. (1995) update genotypes of blocks of individuals jointly at several loci. Jensen and Kong (1999) update arbitrary collections of the latent variables in the pedigree, selected using the HUGIN Bayesian expert system software (Andersen et al., 1989). Heath (1997) and Thompson and Heath (1999) use the meiosis indicators $\mathbf{S} = \{S_{i,j}\}$. Heath (1997) updates jointly the components of $S_{\bullet,j}$, the indicators at a single locus j: the *L-sampler*. Thompson and Heath (1999) update jointly the components of $S_{i,\bullet}$, the meiosis indicators for all loci in a single meiosis i: the *M-sampler*. All these MCMC methods provide, directly or indirectly, realizations of the descent of genes in pedigrees and the genotypes of individuals, and hence Monte Carlo estimates of likelihoods for linkage and segregation analysis (sections 6.2, 6.3 and 7.6), and the probabilities of gene identity by descent and haplotype sharing conditional on observed trait and marker data $\mathbf{Y}$ (section 3.6).

In a Bayesian framework, the segregation and linkage parameters of genetic models are assigned prior probability distributions (see section 2.4). In this case, the same MCMC methods provide estimates of the posterior probability distributions of linkage and trait gene effects and locations.

In the locus-by-locus sampler (L-sampler) first developed by Kong (1991), all genotypes $G_{\bullet,j} = \{G_{i,j}\}$ at a single locus j are updated conditionally upon those at neighboring loci. Computationally the approach is analogous to the sequential imputation method of section 7.5, except that sampling is from the full conditional of $G_{\bullet,j}$. Heath (1997) has further developed the L-sampler, and widened its scope, using $S_{\bullet,j}$ rather than $G_{\bullet,j}$. Because of the structure, this full conditional distribution of $S_{\bullet,j}$ given the data $\mathbf{Y}$ and the meiosis indicators $\mathbf{S}_{-j} = \{S_{\bullet,l}, l \neq j\}$ is

$$P_\theta(S_{\bullet,j} \mid \mathbf{S}_{-j}, \mathbf{Y}) \;=\; P_\theta(S_{\bullet,j} \mid S_{\bullet,j-1}, S_{\bullet,j+1}, Y_{\bullet,j}).$$

That is, the distribution depends only on current values of $S_{\bullet,j-1}$ and $S_{\bullet,j+1}$ and data $Y_{\bullet,j}$. Thus, the calculation of $P_\theta(S_{\bullet,j} \mid \mathbf{S}_{-j}, \mathbf{Y})$ is a single-locus peeling computation analogous to those of section 6.3, and is often feasible. The developments of Heath (1997) are in the context of Bayesian analyses of quantitative traits, under models of several loci contributing additively to the trait value. His approach uses a variety of improved sampling and computational ideas, including more efficient peeling algorithms, integrated proposal distributions (Besag et al., 1995) and reversible jump MCMC (Green, 1995). The output consists of realizations of putative trait loci from a Bayesian posterior; no likelihood or lod score is obtained. One great advantage of the L-sampler is that it is irreducible, provided only that recombination probabilities between adjacent loci are strictly positive. Moreover, this MCMC sampling is a great improvement over single-site methods. However, when there are multiple tightly linked marker loci, mixing can be poor.

8.4 Tightly-linked loci: the M-sampler

The single-site ($S_{i,j}$) or single-locus ($S_{\bullet,j}$) update has mixing problems when loci are tightly linked. An alternative form of block-updating is to update jointly the meiosis indicators for all loci in a given meiosis ($S_{i,\bullet}$). The M-sampler is a whole-meiosis Gibbs sampler (Thompson and Heath, 1999) for $S_{i,\bullet}$. At each step a random meiosis is selected for updating; alternatives in which meioses are updated sequentially are also possible. Note also that, for an unobserved founder with only one offspring in the pedigree, the meiosis from the founder parent to the offspring can be ignored (and not sampled), since there is no information on the haplotypes transmitted.

To implement the M-sampler we must compute

$$\Pr(S_{i,\bullet} \mid \{S_{k,\bullet}, k \neq i\}, \mathbf{Y}).$$

As previously (section 6.2), we suppose that the marker data $\mathbf{Y}$ can be partitioned into data relating to each locus $j = 1, 2, ..., L$, and that the loci are numbered in

order along the chromosome. Then

$$\mathbf{Y} \quad = \quad (Y_{\bullet,1}, ..., Y_{\bullet,L}).$$

As in section 6.2, let

$$Y^{(j)} \quad = \quad (Y_{\bullet,1}, ..., Y_{\bullet,j}), \quad \text{so} \quad \mathbf{Y} = Y^{(L)}.$$

We have seen in section 3.6 that $\Pr(Y_{\bullet,j} \mid S_{\bullet,j})$ can be easily computed.

Now define

$$Q_j^\dagger(s) \quad = \quad \Pr(S_{i,j} = s \mid \{S_{k,\bullet}, k \neq i\},\ Y^{(j)})$$

for $s = 0, 1$. Note that this function $Q_j^\dagger(\cdot)$ is analogous, but not identical, to the function $Q_j^\dagger(\cdot)$ of section 7.1. There the probability considered was the joint distribution for all components of $S_{\bullet,j}$, conditional on $Y^{(j)}$; here the probability is for $S_{i,j}$ conditioning additionally on indicators at other meioses $\{S_{k,\bullet}, k \neq i\}$. Meiosis indicators $S_{i,\bullet}$ are *a priori* independent over i, and become dependent only through conditioning on the data $\mathbf{Y}$ (Figure 8.1). Thus, $Q_j^\dagger(s)$ is the probability for the meiosis indicator $S_{i,j}$, given the data $Y^{(j)}$ and other $(k \neq i)$ meiosis indicators at loci up to and including locus j. (The components $S_{k,l}$ for $l > j$ are irrelevant, since $Y_{\bullet,l}$ is not conditioned upon.) Thus, by analogy with section 7.1, $Q_j^\dagger(s)$ may be computed sequentially just as in equation (7.2). The only difference is that now, rather than considering all 2^m possible values of $S_{\bullet,j}$, we consider only values of the single binary indicator $S_{i,j}$, conditioning on the remainder $(k \neq i)$ which remain fixed. In meiosis i, there is no recombination between locus $(j-1)$ and locus j if the value $(s = 0,\ 1)$ of $S_{i,j}$ is the same as at locus $(j-1)$, and there is recombination if the values differ. That is

$$Q_1^\dagger(s) \quad \propto \quad \Pr(Y_{\bullet,1} \mid S_{\bullet,1})$$

and

$$\begin{aligned} Q_j^\dagger(s) \quad \propto \quad & \Pr(Y_{\bullet,j} \mid S_{\bullet,j})\ (Q_{j-1}^\dagger(s)(1-\rho_{j-1}) \\ & + \ Q_{j-1}^\dagger(1-s)\rho_{j-1}) \end{aligned} \tag{8.8}$$

for $j = 2, ..., L$. In this equation, $S_{\bullet,j}$ takes the current value at meioses k other than i, and the value s for meiosis i. As before, ρ_{j-1} is the recombination frequency between locus $j-1$ and locus j. Thus we may compute (8.8) for each j in turn, working forwards sequentially along the chromosome.

Finally we have computed

$$Q_L^\dagger(s) \quad = \quad \Pr(S_{i,L} = s \mid \{S_{k,\bullet}, k \neq i\},\ \mathbf{Y} = Y^{(L)})$$

and thus $S_{i,L}$ may be sampled from this desired conditional distribution. Suppose now each $S_{i,l}$ has been successively sampled from the required distribution for $l = L, L-1, \ldots, j+1, j$. Then

$$\begin{aligned} & \Pr(S_{i,j-1} = s \mid \{S_{k,\bullet}, k \neq i\}, \{S_{i,l}, l = j, ..., L\}, \mathbf{Y}) \\ & \qquad \propto Q_{j-1}^\dagger(s)\ (T_j\ \rho_{j-1} \ + \ (1-T_j)(1-\rho_{j-1})) \end{aligned} \tag{8.9}$$

where $T_j = |S_{i,j} - s|$ is the indicator of recombination in the interval $j-1$ to j. Thus we may work backwards along the chromosome, sampling each $S_{i,j}$ in turn $(j = L, ..., 1)$, obtaining overall a joint realization of $S_{i,j}$, $j = 1, ..., L$ from its full conditional distribution given $\{S_{k,\bullet}, k \neq i\}$ and $\mathbf{Y}$. Again, this is directly analogous to equation (7.6) of section 7.1.

Throughout this chapter we have ignored the fact that genetic maps differ between males and females: the order of loci is the same, but the recombination frequencies can differ quite widely. Linkage analysis computations should accommodate different values of recombination frequencies for males and females. For the M-sampler this is particularly straightforward, since each meiosis is in a male or in a female. As will be shown in section 11.2, the M-sampler can also incorporate more general meiosis models, including genetic interference, by using a Metropolis-Hastings acceptance/rejection step (Thompson, 2000*a*).

Implementations of almost all the computational algorithms referred to in this chapter are freely available by ftp. The Rockefeller Genetic Linkage Software list at http://linkage.rockefeller.edu/soft/list.html is an excellent reference. The software of our group is implemented primarily in our MORGAN package, which is available by ftp at . The most recent release of MORGAN (MORGAN_VF1, shortly to be replaced by MORGAN_V2.3) includes L-sampler and M-sampler implementations. The site www.stat.washington.edu/thompson/Genepi/pangaea.shtml also includes the *Loki* package for MCMC linkage analysis of quantitative traits (Heath, 1997).