Chapter 7

Monte Carlo Estimates on Pedigrees

7.1 Baum algorithm for conditional probabilities

While the above method of likelihood computation was known to Baum (1972), his primary aim was estimation of the transition probabilities of the Markov chain, and of the probability relationship between input and output (Baum and Petrie, 1966; Baum et al., 1970). Here, these are transition probabilities $P(S_{\bullet,j+1} = s \mid S_{\bullet,j} = s^*)$ and penetrance probabilities $P(Y_{\bullet,j} \mid S_{\bullet,j})$. If the latent variables $\mathbf{S}$ were observed, the sufficient statistics for estimation of these transition and penetrance parameters would be simple functions of $\mathbf{Y}$ and $\mathbf{S}$. Thus, to estimate parameters of the model, for example by using an EM algorithm (Dempster et al., 1977), one must impute these functions of the underlying $\mathbf{S}$ conditional on $\mathbf{Y}$. Again, here we use the notation of meiosis indicators of section 4.7, but the framework is general to any hidden Markov model.

Thus, the forward-backward algorithms of Baum et al. (1970) address *inter alia* the computation of marginal probabilities

$$Q_j(s) = \Pr(S_{\bullet,j} = s \mid \mathbf{Y}),\ j = 1, \ldots, L.$$

We define two functions

$$\begin{aligned} Q_j^\dagger(s) &= \Pr(S_{\bullet,j} = s \mid Y^{(j)}) \\ Q_{j+1}^*(s) &= \Pr(S_{\bullet,j+1} = s \mid Y^{(j)}). \end{aligned}$$

The function $Q_j^\dagger(\cdot)$ provides the imputation of $S_{\bullet,j}$ given data $Y^{(j)}$ up to and including locus j, while $Q_{j+1}^*(\cdot)$ is the predictor of $S_{\bullet,j+1}$ also given $Y^{(j)} = (Y_{\bullet,1}, \ldots, Y_{\bullet,j})$.

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