Chapter 9

Likelihood Ratios for Genetic Analysis

9.1 Monte Carlo likelihood ratio estimation

The MCMC methods of Chapter 8 provide methods for obtaining realizations from $P_\theta(\mathbf{X} \mid \mathbf{Y})$, the probability distribution of latent variables $\mathbf{X}$ conditional on data $\mathbf{Y}$ under a model indexed by parameters θ. In this chapter, we discuss methods of using such realizations in Monte Carlo methods for linkage and segregation analysis, focusing on likelihood methods.

Recall again (equation (7.8)) that, for phenotypic data $\mathbf{Y}$,

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

where latent variables $\mathbf{X}$ are genotypes $\mathbf{G}$ or meiosis indicators $\mathbf{S}$. We again use θ to denote the general set of parameters of a genetic model. These include the recombination or gene location parameters. From equation (7.12), efficient Monte Carlo estimation of $L(\theta)$ will result from sampling from a distribution $P^*(\mathbf{X})$ close to proportional to the joint probability $P_\theta(\mathbf{X}, \mathbf{Y})$:

$$P^*(\mathbf{X}) \;\approx\; P_\theta(\mathbf{X} \mid \mathbf{Y}) \;\propto\; P_\theta(\mathbf{X}, \mathbf{Y}).$$

One possible choice is thus to simulate, by the methods of Chapter 8, not from $P_\theta(\mathbf{X} \mid \mathbf{Y})$ but from $P_{\theta_0}(\mathbf{X} \mid \mathbf{Y})$, where $\theta_0 \approx \theta$. Then

$$\begin{aligned} P_\theta(\mathbf{Y}) \;&=\; \sum_{\mathbf{X}} P_\theta(\mathbf{X}, \mathbf{Y}) \;=\; \sum_{\mathbf{X}} \frac{P_\theta(\mathbf{X}, \mathbf{Y})}{P_{\theta_0}(\mathbf{X} \mid \mathbf{Y})} P_{\theta_0}(\mathbf{X} \mid \mathbf{Y}) \\ &=\; \mathrm{E}_{\theta_0}\left(\frac{P_\theta(\mathbf{X},\, \mathbf{Y})}{P_{\theta_0}(\mathbf{X} \mid \mathbf{Y})} \;\middle|\; \mathbf{Y} \right) \;=\; P_{\theta_0}(\mathbf{Y})\, \mathrm{E}_{\theta_0}\left(\frac{P_\theta(\mathbf{X}, \mathbf{Y})}{P_{\theta_0}(\mathbf{X}, \mathbf{Y})} \;\middle|\; \mathbf{Y} \right). \end{aligned}$$

Hence in genetic analysis, or in any missing-data context, we have the key formula