CHAPTER 5

Nonparametric Maximum Likelihood

We now return to the nonparametric maximum likelihood problem that was introduced in Section 1.6 of Chapter 1, and do the necessary theory to prove the results given there.

The problem is to maximize the mixture likelihood

$$L(Q) = \prod_{i=1}^{n} L_i(Q) = \prod_{j=1}^{D} \left[\int L_j(\phi)\, dQ(\phi) \right]^{n_j}. \tag{5.1}$$

Here $L_j(\phi)$ is the *likelihood kernel*, generally the one-component likelihood for a single observation, say y_j, and n_j is the number of times y_j was observed. The likelihood kernel may well depend on other auxiliary parameters and covariates, which will be held fixed in this discussion. As far as the maximization problem is concerned, the only critical assumption is that L_j is a nonnegative function of ϕ and that the number D is minimal among all such product representations. That is, the terms have been grouped to the maximal extent. In the multinomial setting, this can substantially reduce the number of terms in the product.

5.1. The optimization framework. The basic results concerning the nonparametric maximum likelihood estimator $\hat{Q}$ have already been outlined in Section 1.6. These results can be derived by putting the problem of likelihood maximization into the formal setting of numerical optimization theory. That is, we view it as a problem of the form: maximize an *objective function* $l(\mathbf{p})$ over the elements $\mathbf{p}$ of a set $\mathbf{P}$. If this is done properly, then the results follow readily from standard optimization results.

5.1.1. *Reformulating the problem.* The key to putting this problem into this framework is to examine (5.1) and recognize that the maximum depends directly on the possible values of the *mixture likelihood vector*

$$\mathbf{L}(Q) = (L_1(Q), L_2(Q), \ldots, L_D(Q))'.$$