CHAPTER 7

Extending the Method

In this chapter we present some simple extensions of the NPMLE theorem that solve problems that are similar, but not identical, in structure. We consider three situations: first, a class of problems in which the unknown latent distribution appears in the likelihood in a ratio form; second, the question of maximizing a mixture likelihood with linear constraints on the latent distribution; third, the problem of estimating the latent distribution with a continuous density function.

7.1. Problems with ratio structure.

We start with a simple example that illustrates a problem in which the unknown latent distribution shows up in the likelihood in a ratio form.

7.1.1. *Example: Size bias.* Suppose that $X_1, \ldots, X_n$ are positive-valued random variables, but they arose from a population that was sampled not randomly, but with probabilities that are proportional to some positive function $w(x)$ of the variable of interest. That is, suppose the underlying distribution of the variable X is G, with density g, but the sampling is from the density proportional to $w(x)g(x)$.

A classic example of this type would be if we were to sample vacationers in a hotel lobby and ask how long they were staying in the hotel. The vacationers who have longer stays are more likely to be included in the sample.

The nonparametric MLE problem is then to find the underlying distribution G given knowledge of the sampling weights $w(x)$. We can write the likelihood, for discrete G, as

$$L(G) = \prod \frac{w(x_i)G(\{x_i\})}{\int w(x)\, dG(x)} = \prod \frac{\int w(x_i)\mathscr{I}[\phi = x_i]\, dG(\phi)}{\int w(x)\, dG(x)}.$$

The question is, how do we maximize such a likelihood, which now has the latent distribution in both numerator and denominator?