Statistics 540, EM Algorithm

The EM Algorithm

Goals ...

Provide an iterative scheme for obtaining maximum likelihood estimates, replacing a hard problem by a sequence of simpler problems.

Context ...

Though most apparent in the context of missing data, it is quite useful in other problems as well. The key is to recognize a situation where, if you had more data, the optimization would be simplified.

Approach ...

By *augmenting* the observed data with some additional random variables, one can often convert a difficult maximum likelihood problem into one which can be solved simply, though requiring iteration. Treat the observed data $\boldsymbol{Y}$ as a function $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{X})$ of a larger set of unobserved *complete* data $\boldsymbol{X}$, in effect treating the density

$$g(y;\theta) = \int_{\mathcal{X}(y)} f(x;\theta)dx$$

The trick is to find the right f so that the resulting maximization is simple, since you will need to iterate the calculation.

Computational Procedure ...

The two steps of the calculation that give the algorithm its name are

1. **Estimate** the sufficient statistics of the complete data $\boldsymbol{X}$ given the observed data $\boldsymbol{Y}$ and current parameter values,
2. **Maximize** the $\boldsymbol{X}$ likelihood associated with these estimated statistics.

Genetics Example ...

Observe for some $0 \le \pi \le 1$ counts

$$\boldsymbol{y} = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34) \sim \text{Mult}(1/2 + \pi/4, 1/4(1-\pi), 1/4(1-\pi), \pi/4)$$

Estimate π by ML is messy. Instead, think of $\boldsymbol{y}$ as a collapsed version ($y_1 = x_0 + x_1$) of

$$\boldsymbol{x} = (x_0, x_1, x_2, x_3, x_4) \sim \text{Mult}(1/2, \pi/4, 1/4(1-\pi), 1/4(1-\pi), \pi/4)$$

Steps:

1. Estimate x_0 and x_1 given $y_1 = 125$ and an estimate $\pi^{(i)}$ implies that $x_0^{(i)} = 125(1/2)/(1/2 + \pi^{(i)}/4)$ and $x_1^{(i)} = 125(\pi^{(i)}/4)/(1/2 + \pi^{(i)}/4)$.
2. Maximize the resulting binomial problem, obtaining $\pi^{(i+1)} = \frac{x_1^{(i)}+34}{x_1^{(i)}+34+18+20}$.

Mixture models ...

Suppose that the observed data Y is a mixture of samples from K populations, but that the mixture indicators Z are unknown. Think of $Z_i = (0, 0, \ldots, 1, \ldots, 0)$ as a K vector with one position one and the rest zero. The complete data is $X = (Y, Z)$. The steps in this case are

1. Estimate the group membership probability for each Y_i given the current parameter estimates.
2. Maximize the resulting likelihood, finding in effect the weighted parameter estimates.

References ...

- Dempster, A. P., N. M. Laird, & D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *JRSS-B*, **39**, 1-38.
- Little, R. J. A. and D. B. Rubin (1987). *Statistical Analysis with Missing Data.* Wiley, New York.
- Tanner, M. A. (1993). *Tools for Statistical Inference.* Springer, New York.

Success ?

Theory shows that the EM algorithm has some very appealing monotonicity properties, improving the likelihood at each iteration. Though often slow to converge, it does get there!