STAT 804: Notes on Lecture 9

EM algorithm

The algorithm can be applied when we have (real or imaginary) missing data. Suppose the data we have is X; some other data we didn't get is Y and $Z = (X, Y)$. It often happens that we can think of a Y we didn't observe in such a way that the likelihood for the whole data set Z would be simple. In that case we can try to maximize the likelihood for X by following a two step algorithm first discussed in detail by Dempster, Laid and Rubin. This algorithm has two steps:

1. The **E** or **Estimation** step. We "estimate" the missing data Y by computing $\mathrm{E}(Y|X)$. Technically, we are supposed to estimate the likelihood function based on Z. Factor the density of Z as
$$f_Z = f_{Y|X} f_X$$
and take logs to get
$$\ell(\theta|Z) = \log(f_{Y|X}) + \ell(\theta|X)$$
We actually estimate the log conditional density (which is a function of θ) by computing
$$\mathrm{E}_{\theta_0}(\log(f_{Y|X})|X)$$
Notice the subscript θ_0 on E. This indicates that you have to know the parameter to compute the conditional expectation. Notice too that there is another θ in the conditional expectation – the log conditional density has a parameter in it.

2. We then maximize our estimate of $\ell(\theta|Z)$ to get a new value θ_1 for θ. Go back to step 1 with this θ_1 replacing θ_0 and iterate.

To get started we need a preliminary estimate. Now look at our problem. In our case the quantity Y is ϵ_{-1}. Rather than work with the log-likelihood directly we work with Y. Our preliminary estimate of Y is 0. We use this value to estimate θ as above getting an estimate θ_0. Then we compute $\mathrm{E}_{\theta_0}(\epsilon_{-1}|X)$ and replace ϵ_{-1} in the log-likelihood above by this conditional expectation. Then iterate. This process of guessing ϵ_{-1} is called backcasting.

Summary

- The log likelihood based on $\epsilon_{-1}, X_0, \ldots, X_{T-1}$ is
$$\frac{-\epsilon_{-1}^2}{2\sigma^2} - (T+1)\log(\sigma) - \frac{1}{2}\sum_0^{T-1}(X_t - \psi X_{t-1} - \psi^2 X_{t-2} - \cdots - \psi^{t+1}\epsilon_{-1})^2$$

- Put $\epsilon_{-1} = 0$ in this formula and estimate ψ by minimizing
$$\sum \hat\epsilon_t^2$$
where
$$\hat\epsilon_t = X_t - \psi X_{t-1} - \psi^2 X_{t-2} - \cdots - \psi^t X_0$$
for $t = 0, \ldots, T-1$.

- Now compute $\mathrm{E}(\epsilon_{-1}|X_0, \ldots, X_{T-1})$ Box, Jenkins and Reinsel presents an algorithm to do so based on the fact that there are actually two MA representations of corresponding to a given covariance function (the invertible one and a non-invertible one). The non-invertible representation is

$$X_t = e_t + \frac{1}{\psi} e_{t+1};$$

this form can be used to carry out the computation of the conditional expection.

- Iterate, re-estimating ψ and recomputing the backcast value of ϵ_{-1} if needed.