Inverse Problems in Statistics
Seminar about Statistics SS07
Chapter 5: Computing the Estimates

Ines Bühlmann

4. Juni 2007

5.4. EM algorithm

The Expectation Maximization (EM) Algorithm resembles the problem of the likelihood function. The difference is that now some data are false or simplistic missing.

5.4.1. Example

Consider an i.i.d. sample $X_1, X_2, \ldots, X_n$ from the exponential distribution with density

$$f(x \mid \Theta) = \Theta^{-1} e^{-x/\Theta} 1_{(0,\inf)}(x)$$

However, only the integer part of the random variables can be observed:
$Y_i = \lfloor X_i \rfloor$. The problem is how to estimate $\Theta > 0$ via maximum likelihood from the sample of $Y's$.
For the current iterate Θ, the following function of Θ' is determined:

$$Q(\Theta' \mid \Theta) = E\left(\sum_{i=1}^n \log f(X_i \mid \Theta') \mid \Theta, y_1, \ldots, y_n\right) = \sum_{i=1}^n Q_i(\Theta' \mid \Theta)$$

To compute the function Q, we denote the conditional density of X given Y under the distribution with parameter Θ by k:

$$k(x \mid y; \Theta) = \frac{d}{dx} P(X \leq x \mid Y = y; \Theta) = \frac{e^{-x/\Theta}}{\Theta(e^{-y/\Theta} - e^{-(y+1)/\Theta}} 1_{[y,y+1)}(x)$$

Therefore,

$$Q_i(\Theta' \mid \Theta) = -\log \Theta' - \frac{y_i + \Theta}{\Theta'} + \frac{1}{\Theta'(e^{1/\Theta} - 1)}$$

So, $Q(\Theta' \mid \Theta) = \sum_{i=1}^n Q_i(\Theta' \mid \Theta)$ is maximized at $\Theta' = \Theta + \overline{y_n} - \frac{1}{(e^{1/\Theta} - 1)}$

5.4.2. Idea

Generally the EM algorithm can be described in two steps:
Expectation-step(E-step) and the *Maximization-step* (M-step).
It's given:

- measure space (X, A, μ_1) carrying the complete data X
- second measure space (Y, B, μ_2) supporting the incomplete random vector Y
- measurable mapping $T : X \to Y$
- The space (X, A, μ_1) carries a family of probability distributions $\{P_\theta : \theta \in \Theta\}$ with densities $f(\cdot \mid \theta)$ respect to μ_1 for some parameter set Θ
- The mapping T induces a family of distributions $\{R_\theta = P_\theta \times T^{-1} : \theta \in \Theta\}$ on (Y, B, μ_2) with densities $g(\cdot \mid \theta)$ with respect to μ_2.

Now define

$$L(\theta \mid y) = \log g(y \mid \theta)$$

the loglikelihood function. We are at the point, to define the two steps.

E-step:
The E-step compute the expectation value of $L(\theta \mid y)$.

$$Q(\theta' \mid \theta) \equiv E(L(\theta \mid y)) = E(\log g(y \mid \theta))$$

In this formula θ is not a parameter of the function, but θ is a constant which was till now the best approximation of the parameters of the density function. Otherwise θ' is the parameter of the function, which we want to maximize.

M-step:
The algorithmic map is defined as the argmax of the loglikelihood of the complete observations, given the observed data and the current parameter value.

$$A(\theta) \equiv argmax_{\theta' \in \Theta} Q(\theta' \mid \theta)$$

So $A(\theta)$ is the better approximation of the parameters of the density function than θ'.

Now we can use this better approximation $A(\theta)$ for E-step and M-step again.

In theory the iterationsstep ends, when $A(\theta)$ is equal to the approximation of the loglikelihood function of the density.
But in practice that can not be reached. For example is one reason that the floating points are only approximated.