EM algorithm

John Paige

3/27/2023

TMA4300 - Part 3

- Methods based on resampling
  - Bootstrap
  - Permutation test
- Expectation Maximization Algorithm

Boostrap

- Very general method to estimate characteristics of estimators
- Uses the observations repeatedly
- Can be parametric or non parametric
- Easy to use for iid data
- Can be used for dependent data but then we must sample in a "reasonable" way

Permutation test

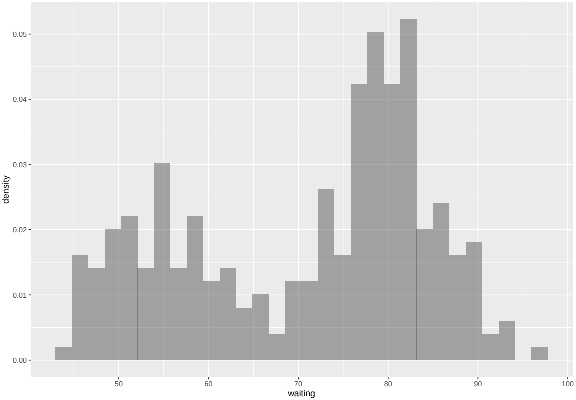
- Non parametric test
- Based on resampling
- Need observations to be exchangeable under the null hypotheses

The expectation-maximization (EM) algorithm

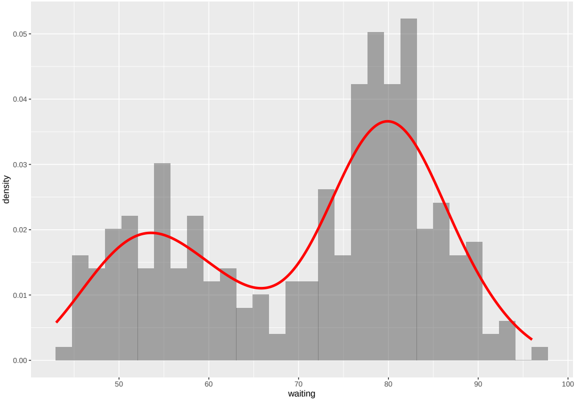
- The expectation maximization (EM) algorithm~(Dempster et al., 2007) is an alternative procedure for the computation of maximum likelihood estimators.
- In certain models – particularly missing data and data augmentation problems – the EM algorithm appears naturally and simplifies the maximum likelihood problem.

Motivation: Old Faithfull Eruptions

Motivation: Old Faithfull Eruptions



Motivation: Old Faithfull Eruptions



Old Faithfull - Time between eruptions

How can we fit this density:

- Non-parametric approach (kernel smoothing, splines)
- Parametric approach (mixtures)

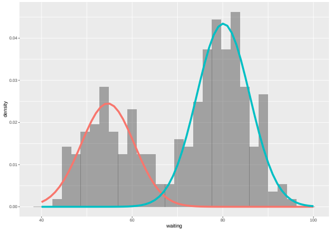
Mixture models:

- Taking a mixture-based approach
  - assume that the sample is made of some (e.g. two) sub-samples (classes)

  -assume that the distribution within each class is known (e.g. normal)
- If this assumptions are reasonable then the red density will give a reasonable fit to the data

$$f(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

```
## number of iterations= 34
```



Statistical model

- We model the data as coming from a normal mixture

$$y_i \sim \pi_1 \mathcal{N}(\mu_1, 1) + \pi_2 \mathcal{N}(\mu_2, 1),$$

for $i = 1, \ldots, n$ with $\pi_1 \in (0, 1)$, $\pi_1 + \pi_2 = 1$

- The vector of unknowns is $\theta = (\pi_1, \mu_1, \mu_2)^\top$
- What is the likelihood function?

See Blackboard

EM for gaussian mixture

1. Take initial guesses for $\hat{\pi}_j$, $\hat{\mu}_j$.
2. E-Step: Compute

$$\hat{\delta}_{z_i}(k) = E\{I(z_i = k)|x_i, \theta\} = \frac{\hat{\pi}_k f_k(x_i|\hat{\mu}_k, 1)}{\sum_{k=1}^{2} \hat{\pi}_k f_k(x_i|\hat{\mu}_k, 1)}$$

3. M-Step: Compute the mixing parameters and weighted means:

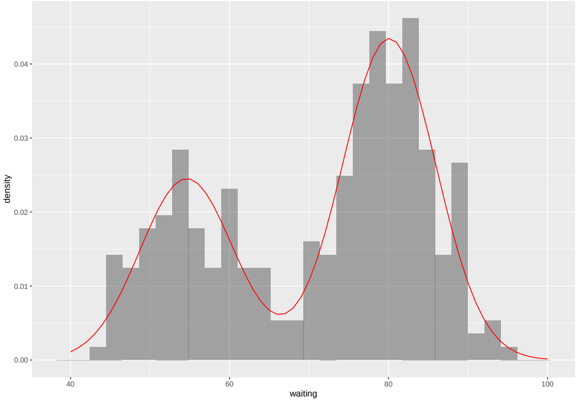
$$\begin{aligned} \hat{\pi}_k &= \quad \frac{\sum_i \delta_{z_i}(k)}{n} \\ \hat{\mu}_k &= \frac{\sum_i \delta_{z_i}(k) x_i}{\sum_i \delta_{z_i}(k)} \end{aligned}$$

4. Iterate steps 2 and 3 until convergence.

EM algorithm for Faithfull data

Results:

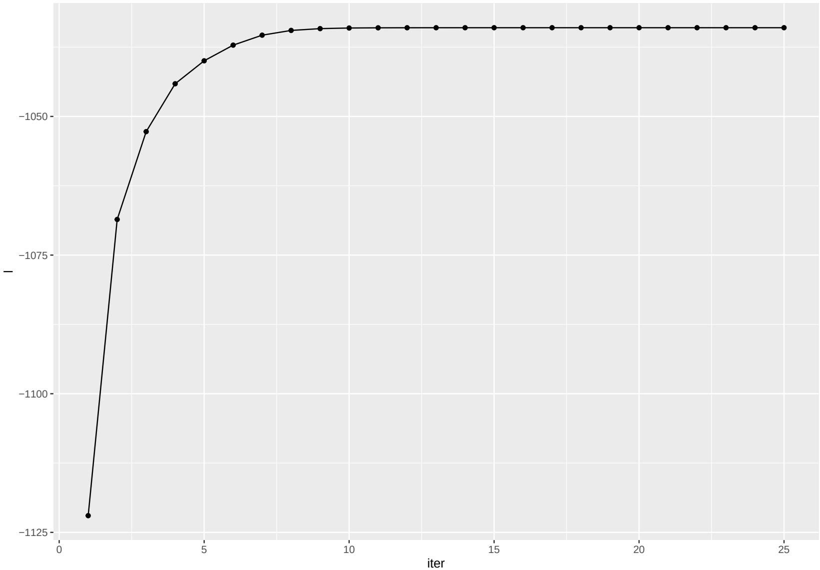
```
## number of iterations= 24
```



$\hat{\pi} = (0.36, 0.64)$, $\hat{\mu} = (54.61, 80.09)$, $\hat{\sigma} = (5.87, 5.87)$

EM algorithm for Faithfull data

Convergence



Direct Maximization

- Possible with the R function `optim()`
- Advantage: Standard errors are directly available
- Important: Choose good starting values

EM algorithm in general

Situation:

- Observed variables X
- Unobserved/hidden variables Z
- Complete dataset $Y = (X, Z)$

The EM algorithm to maximize $\log L(\theta; x)$

We would like to maximize $L(\theta; x)$ regarding θ, but we use $L(\theta; y)$ or rather $\ell(\theta; y) = \log L(\theta; y)$.

Input : Function $\ell(\theta; y)$ and starting value $\theta^{(0)}$

$i \leftarrow 0$
while *not converged* **do**
  E-Step: Compute the conditional expectation

$$Q(\theta) = Q(\theta|\theta^{(i)}) = E(\ell(\theta; y)|x, \theta^{(i)}),$$

  where $\ell(\theta; y)$ is the complete data loglikelihood
  M-Step: Determine

$$\theta^{(i+1)} = \operatorname{argmax}_{\theta \in \Theta} Q(\theta)$$

  Update iteration number: $i \leftarrow i + 1$
end

Gene Example

See Blackboard

Properties of the EM algorithm

- In each iteration step of the EM algorithm the (incomplete) likelihood is increased:

$$L(\theta^{(i+1)}; y) \geq L(\theta^{(i)}; y)$$

- Parameter restrictions are (mostly) automatically fulfilled
- Convergence can be very slow – this especially depends on the "amount' ' of missing data
- Standard errors are not directly available. Some methods exist to try to approximate it, but they are not so easy to use in practice. Much easier just to bootstrap your data!

Frequent applications of the EM algorithm

- Mixture models
- Cluster analysis
- Hidden Markov models
- Parameter estimation with missing data