What to do today (Nov 18, 2020)?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

3. **Essential Topics in Mathematical Statistics**

   3.1 Elementary Statistical Inferences (Chp 4)

   3.2 Consistency and Limiting Distributions (Chp 5)

   3.3 Maximum Likelihood Methods (Chp 6)

   - *3.3.1 Maximum Likelihood Estimation*
   - *3.3.2 Likelihood-Based Tests*
   - **3.3.3 EM Algorithm**

4. *Further Topics, Selected from Chp 7-11*

3.3.3 Expectation-Maximization (EM) Algorithm

What does it do?
An iterative procedure (algorithm) to calculate MLEs of the population parameters $\boldsymbol{\theta}$ when it is hard to maximize the likelihood function of θ with the available data.

Original References:

- Dempster, Laird, and Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. 39 (1): 1–38.
- C.F. Jeff Wu (1983). "On the Convergence Properties of the EM Algorithm". *Annals of Statistics*. 11 (1): 95–103.
- B.W. Turbull (1976) "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data". *Journal of the Royal Statistical Society, Series B*. 38 (3): 290–295.

3.3.3 Expectation-Maximization (EM) Algorithm

Goal: Consider to maximize $L(\boldsymbol{\theta}|\mathbf{X})$, the (observed) likelihood function with the available data $\mathbf{X}$ to obtain the MLE of population parameter $\boldsymbol{\theta}$.
Suppose that it is relatively easier to maximize $L_C(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})$, the (complete) likelihood function of $\boldsymbol{\theta}$ with the "augmented" data $(\mathbf{X}, \mathbf{Z})$.

Define $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{x}) = E_{\boldsymbol{\theta}_0}\left[\log L_C(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})|\boldsymbol{\theta}_0, \mathbf{X} = \mathbf{x}\right]$.

EM Algorithm. Let $\hat{\boldsymbol{\theta}}^{(m)}$ be the estimate on the mth step with $m \geq 0$. The following compute the estimate on the $(m+1)$th step:

- *Expectation-Step.* Compute

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}, \mathbf{x}) = E_{\hat{\boldsymbol{\theta}}^{(m)}}\left[\log L_C(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})|\hat{\boldsymbol{\theta}}^{(m)}, \mathbf{X} = \mathbf{x}\right].$$

- *Maximization-Step.* Let the updated estimate be

$$\hat{\boldsymbol{\theta}}^{(m+1)} = argmax_{\boldsymbol{\theta}\in\Omega} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}, \mathbf{x}).$$

3.3.3 Expectation-Maximization (EM) Algorithm

- Starting with an initial estimate (guess) of $\boldsymbol{\theta}$, say, $\hat{\boldsymbol{\theta}}^{(0)}$, repeat the *E-Step* and *M-Step* and record the sequence of estimates $\{\hat{\boldsymbol{\theta}}^{(m)} : m = 1, 2, \dots\}$.

- Under some assumptions, the sequence converges in probability to the MLE $\widehat{\boldsymbol{\theta}}$ as $m \to \infty$.

- The EM algorithm works to improve $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}, \mathbf{x})$. This implies improvements to $L(\boldsymbol{\theta}|\mathbf{x})$: $L(\hat{\boldsymbol{\theta}}^{(m+1)}|\mathbf{x}) \geq L(\hat{\boldsymbol{\theta}}^{(m)}|\mathbf{x})$.

EM algorithm has a broad range of applications!

... and a lot of variants, eg MCEM Algorithm, ES Algorithm, ...

Example 3.14 Consider a mixture of normal distributions: $X = (1-W)Y_1 + WY_2$ with $Y_j \sim N(\mu_j, \sigma_j^2)$ for $j = 1, 2$ and $W \sim B(1, \epsilon)$. Suppose the observations on a random sample $\mathbf{X}' = (X_1, \ldots, X_n)$ from the mixture distn are available.

- Apply the EM algorithm to estimate the parameter $\boldsymbol{\theta}' = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \epsilon)$
- Similar to Ex6.6.8, generate data with $n = 1000$ and $\boldsymbol{\theta}' = (100, 15^2, 200, 10^2, .8)$, and then evaluate the MLE of $\boldsymbol{\theta}'$ by EM algorithm.

Solution. Let $f_j(y)$ be the pdf of Y_j. The log-likelihood functn with the observed data is

$$l(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log\left[(1-\epsilon)f_1(x_i) + \epsilon f_2(x_i)\right].$$

With observations $\mathbf{w}'$ on a random sample $W_1, \ldots, W_n$ from $W \sim B(1, \epsilon)$, the log-likelihood functn with the complete data is

$$l_C(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n} \left\{ \left[(1-w_i)\log f_1(x_i) + w_i \log f_2(x_i)\right] + \left[w_i \log \epsilon + (1-w_i)\log(1-\epsilon)\right] \right\}.$$

Provided the estimate at the mth stage is $\boldsymbol{\theta}^{(m)}$,

E-Step. $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}, \mathbf{x}) = l_C(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w})\big|_{\mathbf{w}=\widehat{\mathbf{w}}^{(m)}}$ with

$$\hat{w}_i^{(m)} = E(W_i|\boldsymbol{\theta}^{(m)}, \mathbf{x}) = \frac{\epsilon^{(m)} f_2(x_i)^{(m)}}{(1-\epsilon^{(m)})f_1(x_i)^{(m)} + \epsilon^{(m)} f_2(x_i)^{(m)}}.$$

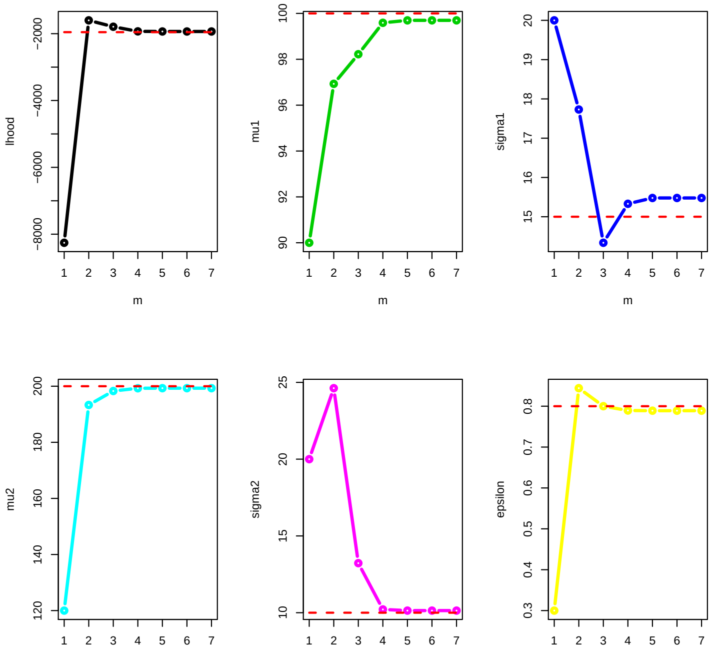
M-Step. Maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}, \mathbf{x})$ wrt $\boldsymbol{\theta}$ yields

$$\mu_1^{(m+1)} = \frac{\sum_{i=1}^n (1-\hat{w}_i^{(m)})x_i}{\sum_{i=1}^n (1-\hat{w}_i^{(m)})}, \quad \sigma_1^{2\,(m+1)} = \frac{\sum_{i=1}^n (1-\hat{w}_i^{(m)})(x_i - \mu_1^{(m+1)})^2}{\sum_{i=1}^n (1-\hat{w}_i^{(m)})},$$

$$\mu_2^{(m+1)} = \frac{\sum_{i=1}^n \hat{w}_i^{(m)} x_i}{\sum_{i=1}^n \hat{w}_i^{(m)}}, \quad \sigma_2^{2\,(m+1)} = \frac{\sum_{i=1}^n \hat{w}_i^{(m)}(x_i - \mu_2^{(m+1)})^2}{\sum_{i=1}^n \hat{w}_i^{(m)}},$$

and $\epsilon^{(m+1)} = \sum_{i=1}^n \hat{w}_i^{(m)} / n$.

$$\boldsymbol{\theta}^{(0)} = (90, 20^2, 120, 20^2, .3);\ m = 7;$$
$$\widehat{\boldsymbol{\theta}} = (99.70, 15.48^2, 199.30, 10.14^2, 0.79)$$



What will we do next week?

1. *Introduction*
2. *Probability and Distribution (Chp 1-3)*
3. *Essential Topics in Mathematical Statistics (Chp 4-6)*
4. **Further Topics, Selected from Chp 7-11**
   - **4.1 Nonparametric and Robust Statistics (Chp 10)**
   - *4.2 Bayesian Procedures (Chp 11.1)*