Chapter 7: EM algrithm and NPMLE: JAW Ch.4 ctd

7.1 The EM Algorithm.

(i) See hard-copy handout notes from Stat 582, 2001.

(ii) Recall homework example (with added parameter σ^2):
Y_i i.i.d. from mixture $\frac{1}{2}N(-\theta,\sigma^2)+\frac{1}{2}N(\theta,\sigma^2)$

(iii)

$$\begin{aligned} f_Y(y;\theta,\sigma^2) &= \frac{1}{2}(2\pi\sigma^2)^{-\frac{1}{2}}\exp(-y^2/(2\sigma^2))\exp(-\theta^2/(2\sigma^2)) \\ &\qquad (\exp(-y\theta/\sigma^2)+\exp(y\theta/\sigma^2)) \\ \ell_n(\theta,\sigma^2) &= \text{const}-(n/2)\log(\sigma^2)-(\sum_i Y_i^2)/(2\sigma^2)-n\theta^2/(2\sigma^2) \\ &\qquad +\sum_i \log((\exp(-Y_i\theta/\sigma^2)+\exp(Y_i\theta/\sigma^2)) \end{aligned}$$

(iv) What is the sufficient statistic?
How would we estimate (θ,σ^2) ?

(v) Now suppose each Y_i caries a "flag" $Z_i = -1$ or 1 as obsn i comes from $N(-\theta,\sigma^2)$ or $N(\theta,\sigma^2)$. Let $X_i=(Y_i,Z_i)$.

$$\begin{aligned} f_X(y,z;\theta,\sigma^2) &= \frac{1}{2}(2\pi\sigma^2)^{-\frac{1}{2}}\exp(-(y-\theta z)^2/2\sigma^2) \\ \ell_{c,n}(\theta,\sigma^2) &= \sum_i \log f(Y_i,Z_i;\theta,\sigma^2) \\ &= \text{const}-(n/2)\log(\sigma^2)-\sum_i (Y_i-\theta Z_i)^2/2\sigma^2 \\ &= \text{const}\ -(n/2)\log(\sigma^2)-(2\sigma^2)^{-1}(\sum Y_i^2-\theta\sum Y_iZ_i+\theta^2 n) \end{aligned}$$

(vi) What now is sufficient statistic, if X_i were observed?
How now could you estimate (θ,σ^2) ?
The Z_i are "latent variables"; X_i are "complete data"

(vii) $\mathrm{E}(\ell_{c,n}|Y)$ requires only

$$\mathrm{E}(Z_i|Y_i) = \frac{\phi((y_i-\theta)/\sigma)-\phi((y_i+\theta)/\sigma)}{\phi((y_i-\theta)/\sigma)+\phi((y_i+\theta)/\sigma)}$$

7.2 Why does EM work – see handout notes

7.3 Types of examples

(i) Multinomial examples – e.g. ABO blood types, see handout

(ii) Mixture examples – see 7.1, also hwk.

(iii) Missing data
Caution: we do NOT "use expectations to impute the missing data"
We compute the expected complete-data log-likelihood. This normally involves using conditional expectations to impute the complete-data sufficient statistics. This is NOT the same thing – see hwk. And it could be more complicated than this – although not if we have chosen sensible "complete-data".

(iv) Censored data, age-of-onset-data, competing risks models, etc.

(v) Hidden states, latent variables:
Models in Genetics, Biology, Climate modelling, Environmental modelling.

(vi) General auxiliary variables: the latent variables do not have to mean anything – they are simply a tool, s.t. that the complete-data log-likelihood is easy.

7.4 Review of Exponential families

(i) See handout notes, and/or earlier notes.
Form, examples – see 2.2

(ii) The natural sufficient statistics. – see 2.4.
The natural parameter space.

(iii) Moment formulae. see 2.2 (vii)

$$\begin{aligned} \mathrm{E}(t_j(X)) &= -\frac{\partial \log c(\pi)}{\partial \pi_j} \\ \mathrm{Cov}(t_j(X), t_j(X)) &= -\frac{\partial^2 \log c(\pi)}{\partial \pi_j \partial \pi_l} \end{aligned}$$

(iv) Likelihood equation for exponential family – see 4.6

$$\begin{aligned} \frac{\partial \ell}{\partial \pi_j} &= n(n^{-1} \sum_1^n t_j(X_i) - \mathrm{E}(t_j(X))) \\ I(\pi) &= J(\pi) \quad = \quad \mathrm{var}(T_1, ..., T_k) \\ I(\tau) &= (\mathrm{var}(T_1, ..., T_k))^{-1} \quad \text{where } \tau_j = \mathrm{E}(t_j(X)) \end{aligned}$$

$(T_1, ..., T_k)$ **achieves (multiparameter) CRLB for** $(\tau_1, ..., \tau_k)$**.**

(v) Completeness see 2.4 (vii)

7.5 EM for exponential families

(i) Suppose complete-data X has exp.fam. form: for n-sample $T_j(X) = \Sigma_{i=1}^{n} t_j(X_i)$

$$\begin{aligned}\log g_\theta(X) &= \log c(\theta) + \sum_{j=1}^{k} \pi_j(\theta) T_j(X) + \log h(X) \\ Q(\theta; \theta^*) &= \log c(\theta) + \sum_{j=1}^{k} \pi_j(\theta) \mathrm{E}_{\theta^*}(T_j(X)|Y) + \mathrm{E}_{\theta^*}(\log h(X)|Y).\end{aligned}$$

(ii) In natural parametrization π_j:

$$\begin{aligned}Q(\pi; \pi^*) &= \log c(\pi) + \sum_{j=1}^{k} \pi_j \mathrm{E}_{\pi^*}(T_j(X)|Y) \\ \frac{\partial Q}{\partial \pi_j} &= \mathrm{E}_{\pi^*}(T_j(X)|Y) + \frac{\partial}{\partial \pi_j} \log c(\pi) \\ &= \mathrm{E}_{\pi^*}(T_j(X)|Y) - \mathrm{E}_{\pi}(T_j(X))\end{aligned}$$

Thus EM iteratively fits unconditioned to conditioned expectations of T_j. At MLE $\mathrm{E}_{\pi^*}(T_j(X)|Y) = \mathrm{E}_{\pi^*}(T_j(X))$.

(iii) Recall

$$\begin{aligned}\ell(\pi) &= \log g_\pi(X) - \log g^*_\pi(X|Y) \\ \text{but} \qquad g^*_\pi(X|Y) &= \frac{h(X)\exp(\Sigma_j \pi_j t_j(X))}{\int_{y(X)=y} h(X)\exp(\Sigma_j \pi_j t_j(X)) dX} \\ &= c^*(\pi; Y) h(X) \exp(\sum_j \pi_j t_j(X)) \\ \text{so} \qquad \ell(\pi) &= \log c(\pi) - \log c^*(\pi; Y)\end{aligned}$$

(iv) Hence, differentiating this:

$$\begin{aligned}\frac{\partial \ell}{\partial \pi_j} &= -\mathrm{E}_\pi(T_j) + \mathrm{E}_\pi(T_j|Y) \\ \text{At MLE}: \qquad \mathrm{E}_\pi(T_j) &= \mathrm{E}_\pi(T_j|Y)\end{aligned}$$

(v) Differentiating again:

$$-\frac{\partial^2 \ell}{\partial \pi_j \partial \pi_l} \;=\; \mathrm{Cov}(T_j, T_l) \;-\; \mathrm{Cov}((T_j, T_l)|Y)$$

If Y determines X, $\mathrm{var}(T(X)|Y) = 0$, and then observed information is $\mathrm{var}(T)$ as for any exp fam.
If Y tells nothing about X, $\mathrm{var}(T(X)|Y) = \mathrm{var}(T(X))$, and observed information is 0.
"Information lost" due to observing Y not X is $\mathrm{var}(T(X)|Y)$.