Goal: get posterior on parameters of posterior:

$$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho \mid X) \propto \left(\prod_{i=1}^{n} P(X_i \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho) \right) P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)$$

5/4/17

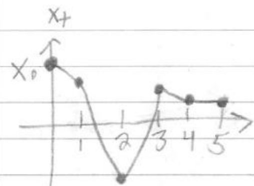
N-R Method

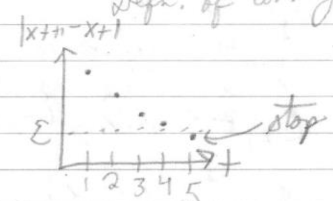
$f(x) = 0$, solve for x. Given ε.

① Guess solution in x_0.

② Calculate $x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$ } iterative steps in an iterative algorithm

③ Repeat step 2 until $\underbrace{|x_{t+1} - x_t| < \varepsilon}_{\text{Defn. of convergence}}$

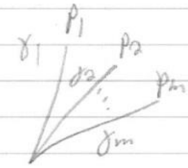




(Likelihood is a mixture) Consider the model

$$X_1, \ldots, X_n \mid \vec{\theta}_1, \ldots, \vec{\theta}_m, \gamma_1, \ldots, \gamma_M \overset{iid}{\sim} \sum_{m=1}^{M} \gamma_m P_m(\vec{\theta}_m)$$

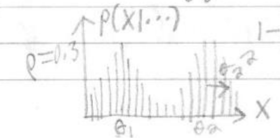
s.t. $\gamma_1 + \gamma_2 + \ldots + \gamma_m = 1$



e.g.

$$X_1, \ldots, X_n \mid \underbrace{\theta_1, \sigma_1^2}_{\vec{\theta}_1}, \underbrace{\theta_2, \sigma_2^2}_{\vec{\theta}_2}, \underbrace{\rho}_{\gamma_1, \gamma_2}$$

$$\overset{iid}{\sim} \underset{\gamma_1}{\rho}\, \overbrace{N(\theta_1, \sigma_1^2)}^{P_1(\vec{\theta}_1)} + \underset{\gamma_2}{(1-\rho)}\, \underbrace{N(\theta_2, \sigma_2^2)}_{P_2(\vec{\theta}_2)}$$



Bayesian Inference: Here, inferring to get parameter values

(likelihood)

$$P(\theta_1,\sigma_1^2,\theta_2,\sigma_2^2,\rho|X) \propto P(X|\theta_1,\sigma_1^2,\theta_2,\sigma_2^2,\rho)\ \underset{(\text{prior})}{\underbrace{P(\theta_1,\sigma_1^2,\theta_2,\sigma_2^2,\rho)}}$$

$$= \begin{array}{ccccc} P(\theta_1) & P(\sigma_1^2) & P(\theta_2) & P(\sigma_2^2) & P(\rho) \\ 1, & 1, & 1, & 1, & 1, \\ & \frac{1}{\sigma_1^2} & & \frac{1}{\sigma_2^2} & \end{array}$$

$$= \left(\prod_{i=1}^{n} \rho \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2} + (1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}\right) \underbrace{\frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}}_{=\text{prior}}$$

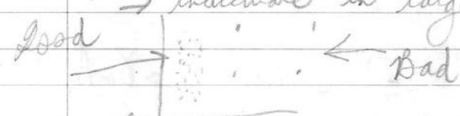
$$= K(\theta_1,\sigma_1^2,\theta_2,\sigma_2^2,\rho|X)$$

<u>How to get inference?</u>

<u>Grid search</u>

$G_{\theta_1} = \langle \ldots \rangle,\ G_{\sigma_1^2} = \langle \ldots \rangle, \ldots$

$\Rightarrow$ inaccurate in large M high dimensions



What if we knew which component each X_i belongs to?

Define:

$I_1 := \mathbb{1}_{X_1}$
$I_2 := \mathbb{1}_{X_2}$
$\vdots$
$I_n := \mathbb{1}_{X_n}$

latent variables / information
$\downarrow$
the I_i's are unobserved but important

Let $I = \{I_1, \ldots, I_n\}$

Recall

$$f(z) = \int f(z,y)\,dy = \int f(z|y) f(y)\,dy$$

$$\underbrace{P(X|\theta) = \int P(X, I|\theta)\,dI}_{\text{This is called "data augmentation"}} = \int P(X|I, \theta)\, P(I, \theta)\, dI$$

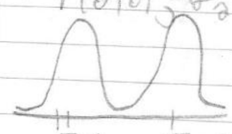
Augment with I's, the latent variable

$$P(\theta, \sigma_1^2, \theta_2, \sigma_2^2, \rho | X) \propto$$

1977 $$\int P(X|I, \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)\, P(I|\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)$$

$$\cdot P(\theta_1 \sigma_1^2, \theta_2^2, \rho)\, dI = k(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho | X)$$

$$= \int k(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho | X, I)\, dI$$



Modest Goal:

Get $\hat{\theta}_{MAP} = \operatorname{argmax}\{K(\theta, X)\}$

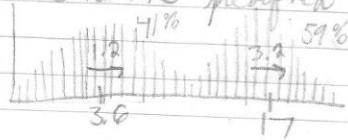
Expectation - Maximization Algorithm (1977)

Step 1: Guess $\hat{\theta}_{MAP} = \theta_0$ to start.

Expectation Step — Step 2: Compare $I_0 = E[I | X, \theta = \theta_0]$

Maximization Step — Step 3: Consider $\mathcal{L}(\theta; I_0, X) = k(\theta | I_0, X)$ (body of integral)
and find $\hat{\theta}_1 = \operatorname{argmax}\{\mathcal{L}(\theta; I_0, X)\}$ i.e. the MLE procedure

Step 4: Repeat steps 2 & 3 until $\|\theta_{t+1} - \theta_t\| < \varepsilon$, where ε is the predefined tolerance level.



E-M Implementation for our 2-Normal Mixture

Step 1: Initialize

$\theta_{1,0} = 0$; $\sigma_{1,0}^2 = 1$, $\theta_{2,0} = 0$, $\sigma_{2,0}^2 = 1$, $\rho = 0.5$

Step 2

$$I_{1,0} = E[I_1 \mid X, \theta_1 = \theta_{1,0}, \sigma_1^2 = \sigma_{1,0}^2, \theta_2 = \theta_{2,0}, \sigma_2^2 = \sigma_{2,0}^2, p = p_0]$$

$$= P(I_1 = 1 \mid X, \dots) = \frac{P(X \mid I_1 = 1, \dots) P(I_1 = 1 \mid \dots)}{\underbrace{P(X \mid \dots)}_{P(X \mid I_1 = 1, \dots) \cdot P(I_1 = 1 \mid \dots) + P(X \mid I_1 = 0, \dots) \cdot P(I_1 = 0 \mid \dots)}}$$

$Q_i = \mathbb{1}_A \sim Bern(P(A))$

$E[Q] = P(A)$

$I_1 \sim Bern(P(I_1 = 1 \mid \dots))$

p: $m=1$, $I = 1$

$1-p$: $m=0$, $I = 0$

E-M Implementation for 2-normal mixture

Step 1: Initialize

$\theta_{1,0} = 0$

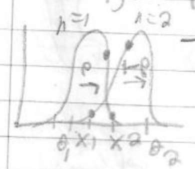
$\sigma_{1,0}^2 = 1$

$\theta_{2,0} = 0$

$\sigma_{2,0}^2 = 1$

$p = 0.5$

Step 2



$$I_{1,0} = \frac{p \frac{1}{\sqrt{2\pi\sigma_{1,0}^2}} e^{-\frac{1}{2\sigma_{1,0}^2}(X_1 - \theta_{1,0})^2}}{p \frac{1}{\sqrt{2\pi\sigma_{1,0}^2}} e^{-\frac{1}{2\sigma_{1,0}^2}(X_1 - \theta_{1,0})^2} + (1-p) \frac{1}{\sqrt{2\pi\sigma_{2,0}^2}} e^{-\frac{1}{2\sigma_{2,0}^2}(X_1 - \theta_{2,0})^2}}$$

$I_{2,0} = E[I_2 \mid X_2 \dots]$

$\vdots$

$\hat{I}_{3,0} = E[I_3 \mid X_3, \dots]$

$\vdots$

$\hat{I}_{n,0} = E[I_n \mid X_n, \dots]$

$\mathcal{L}(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, p; I, X) = \mathcal{L}$

$(P(X \mid I, \theta) P(I \mid \theta) P(\theta))$

$$\mathcal{L} = \left(\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2} \right)^{I_i} \right.$$

$$\cdot \left. \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2} \right)^{1-I_i} \right) \left(\prod_{i=1}^{n} \rho^{I_i} (1-\rho)^{1-I_i} \right)$$

$$\cdot \left((\sigma_1^2)^{-1} (\sigma_2^2)^{-1} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}} \right)^n (\sigma_1^2)^{-1} (\sigma_2^2)^{-1} (\sigma_1^2)^{-\frac{1}{2}\sum I_i} (\sigma_2^2)^{\sum 1-I_i}$$

$$\cdot\, e^{-\frac{1}{2\sigma_1^2} \sum I_i (X_i-\theta_1)^2 - \frac{1}{2\sigma_2^2} \sum (1-I_i)(X_i-\theta_2)^2}$$

$$\cdot\, \rho^{\sum I_i} (1-\rho)^{\sum 1-I_i}$$

$$\ell(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho; I, X) = n \ln\left(\frac{1}{\sqrt{2\pi}}\right)$$

$$-\left(1+\tfrac{1}{2}\sum I_i\right)\ln(\sigma_1^2) - \left(1+\tfrac{1}{2}\sum 1-I_i\right)\ln(\sigma_2^2)$$

$$- \frac{1}{2\sigma_1^2} \sum I_i (X_i-\theta_1)^2 - \frac{1}{2\sigma_2^2} \sum (1-I_i)(X_i-\theta_2)^2$$

$$+ \left(\sum I_i\right) \ln \rho + \Big(\underbrace{\sum 1-I_i}_{n-\sum I_i}\Big) \ln(1-\rho).$$

$$\rightarrow -\left(\frac{2 I_i X_i^2}{2\sigma_1^2} - \frac{\theta_1 \sum x_i I_i}{\sigma_1^2} + \frac{\theta_1^2 \sum I_i}{\sigma_1^2} \right)$$

$$-\left(\frac{\sum (1-I_i) X_i^2}{2\sigma_2^2} - \frac{\theta_2 \sum X (1-I_i)}{\sigma_2^2} + \frac{\theta_2^2 \sum I_i}{2\sigma_2^2} \right)$$

Get $\hat{\theta}_1$ by $\frac{\partial}{\partial \theta_1}[\ell] \overset{set}{=} 0$

$$\frac{\sum x_i I_i}{\sigma_1^2} - \frac{2\theta_1 \sum I_i}{2\sigma_1^2} = 0$$

$$\Rightarrow \hat{\theta}_1 = \frac{\sum x_i I_i}{\sum I_i} \qquad \overline{X}_{mix\,1}$$

$\hat{\theta}_2 = \dfrac{\Sigma X_i(1-I_i)}{\Sigma 1-I_i}$, $\bar{x}_{mix\,2}$

Get $\hat{\sigma}_1^2$

$\dfrac{\partial}{\partial \sigma_1^2}[l] = -\dfrac{1+\frac{1}{2}\Sigma I_i}{\cancel{\sigma_1^2}} + \dfrac{\Sigma I_i(X_i-\theta)^2}{2(\sigma_1^2)^2} = 0$

$\Rightarrow 1+\frac{1}{2}\Sigma I_i = \dfrac{1}{2\sigma_1^2}\Sigma I_i(X_i-\theta_1)^2$

$\Rightarrow 2+\Sigma I_i = \dfrac{1}{\sigma_1^2}\Sigma I_i(X_i-\theta_1)^2$

$\Rightarrow \hat{\sigma}_1^2 = \dfrac{\Sigma I_i(X_i-\theta)^2}{2+\Sigma I_i}$

Likewise, $\hat{\sigma}_2^2 = \dfrac{\Sigma(1-I_i)(X_i-\theta)^2}{2+\Sigma(1-I_i)}$

$\hat{\rho}$ get $\dfrac{\partial}{\partial\rho}[l] \stackrel{set}{=} 0$

$= \dfrac{\Sigma I_i}{\rho} - \dfrac{\Sigma 1-I_i}{1-\rho} = 0$

Use data until converges

$\dfrac{\Sigma I_i}{\rho} = \dfrac{\Sigma 1-I_i}{1-\rho} = n-\Sigma I_i$

Gibbs

$\Sigma I_i - \rho\Sigma I_i = \rho n - \cancel{\rho\Sigma I_i} \Rightarrow \boxed{\hat{\rho} = \dfrac{\Sigma I_i}{n}}$