$$\pi_i^{(t)} = -\log\left\{1 + \frac{1-\lambda}{\lambda}\,\frac{\phi(x_i;\mu_2,\Sigma_2)}{\phi(x_i;\mu_1,\Sigma_1)}\right\}$$

$$= -\log\left(1 + \exp\left\{\log\frac{1-\lambda}{\lambda} + \log\phi(x_i;\mu_2,\Sigma_2) - \log\phi(x_i;\mu_1,\Sigma_1)\right\}\right)$$

Takeaway: never work with densities

always work with $\log f$

EM Convergence

Since $\mathbb{P}(x,z;\theta) = \mathbb{P}(z|x;\theta)\,\mathbb{P}(x;\theta)$

$$\log\mathbb{P}(x;\theta) = \log\mathbb{P}(x,z;\theta) - \log\mathbb{P}(z|x;\theta)$$

$$\mathbb{E}_{z|x,\theta}\left[\log\mathbb{P}(x;\theta)\right]$$

$$= \mathbb{E}_{z|x}\left[\log \mathbb{P}(x, z; \theta)\right] - H_{z|x}\left[\log \mathbb{P}(z|x)\right]$$

$$\log \mathbb{P}(x;\theta) = Q(\theta;\theta^{(t)}) - \underbrace{H(\theta;\theta^{(t)})}_{\text{cross entropy}}$$

$$\ell(\theta;x) = Q(\theta;\theta^{(t)}) - H(\theta;\theta^{(t)})$$

We can show that

$H(\theta;\theta^{(t)})$ is maximal at $\theta^{(t)}$

$$H(\theta^{(t)};\theta^{(t)}) - H(\theta;\theta^{(t)})$$

$$= \underbrace{D\left\{\mathbb{P}(z|x;\theta^{(t)}) \,\|\, \mathbb{P}(z|x;\theta)\right\}}_{\text{KL Divergence}} \geq 0$$

Thus

$$\ell(\theta^{(t+1)};\theta) - \ell(\theta^{(t)};x)$$

$$= \left\{ Q(\theta^{(t+1)}; \theta^{(t)}) - Q(\theta^{(t)}; \theta^{(t)}) \right\}$$

$$- \left\{ \underbrace{H(\theta^{(t)}, \theta^{(t)}) - H(\theta^{(t+1)}; \theta^{(t)})}_{\geq 0} \right\}$$

From the M step we know

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta; \theta^{(t)})$$

So

$$Q(\theta^{(t+1)}; \theta^{(t)}) - Q(\theta^{(t)}; \theta^{(t)}) \geq 0$$

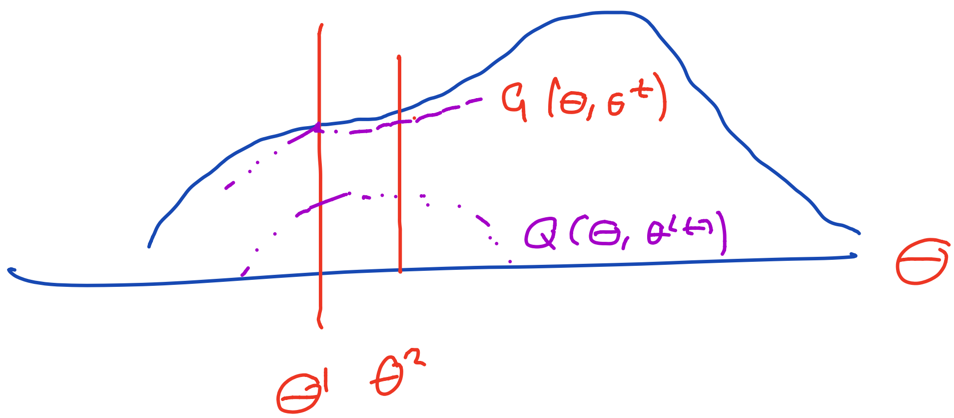
Hence

$$\ell(\theta^{(t+1)}; \theta^{(t)}) \geq \ell(\theta^{(t)}; \theta^{(t)})$$

Note that

$$\ell(\theta; x) \geq \ell(\theta^{(t)}; x) + Q(\theta; \theta^{(t)}) - Q(\text{[illegible]}$$

$$L(\theta, X) = \underbrace{\ell(\theta^{(t)} : X) + Q(\theta : \theta^{(t)}) - Q(\theta^{t}; \theta^{t})}_{= G(\theta, \theta^{(t)})}$$



* optimizes a lower bound on $\ell(\theta)$

Relaxation: We really just needed cases where

$$Q(\theta^{(t+1)}, \theta^{(t)}) - Q(\theta^{t}; \theta^{(t)}) \geq 0$$

This more general M step is called <u>Generalized EM</u> (GEM)

Eg: "Expectation – Conditional – Max"

(ECM)

1. E-step as usual

$$Q(\theta, \theta^{(t)}) = E_{z \mid x, \theta^{(t)}}\left[\ell(\theta; x, z)\right]$$

2. M-step: Conditional Maximization

If $|\theta| = n$ then perform n substeps

$$\theta_1^{(t+1)} = \underset{\theta_1}{\operatorname{argmax}}\ Q\left(\theta_1, \theta_2^{(t)}, \ldots, \theta_n^{(t)}\right)$$

$$\theta_2^{(t+1)} = \underset{\theta_2}{\operatorname{argmax}}\ Q\left(\theta_1^{(t+1)}, \theta_2, \ldots, \theta_n^{(t)}\right)$$

$$\vdots$$

$$\theta_n^{(t+1)} = \underset{\theta_n}{\operatorname{argmax}}\ Q\left(\theta_1^{(t+1)}, \ldots, \theta_n\right)$$

Note that

$$
\begin{aligned}
Q\left(\theta^{(t+1)}; \theta^{(t)}\right) &\geq Q\left(\theta_1^{(t+1)}, \ldots, \theta_n^{(t)}\right) \\
&\geq \\
&\ \ \vdots \\
&\geq Q\left(\theta^{(t)}; \theta^{(t)}\right)
\end{aligned}
$$