1a) SUPPOSE TO HAVE A SET OF 1-DIMENTIONAL POINTS $x_i \in \mathbb{R}$

$X = (x_1, ..., x_n)$ THAT ARE UNDER THE PROBABILITY DISTRIBUTION

$$P(x,\sigma) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & x \geq 0 \\ 0 & x < 0 \end{cases} = \{2\sigma^2, \lambda\} = \begin{cases} \frac{2x}{\lambda} e^{-\frac{x^2}{\lambda}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

WHAT WILL BE THE M.L.E. (λ)?

$$L = \prod_{n=1}^{N} P(x_n|\lambda) \rightarrow \tilde{L} = \ln\left(\prod_{n=1}^{N} P(x_n|\lambda)\right) = \sum_{n=1}^{N} \ln\left(\frac{2x_n}{\lambda} e^{-\frac{x_n^2}{\lambda}}\right) =$$

$$= \sum_{n=1}^{N} \left(\ln(2x_n) - \ln(\lambda) - \frac{x_n^2}{\lambda}\right)$$

$$\frac{\partial \tilde{L}}{\partial \lambda} = \sum_{n=1}^{N} \left(-\frac{1}{\lambda} + \frac{x_n^2}{\lambda^2}\right) = 0 \qquad -\frac{N}{\lambda} + \frac{\sum_{n=1}^{N} x_n^2}{\lambda^2} = 0 \qquad \lambda = \frac{\sum_{n=1}^{N} x_n^2}{N}$$

1b) IN ORDER TO MAKE λ MORE APPROPRIATE, WE CAN IMPOSE THAT $P(\lambda, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\frac{\beta}{\lambda}}$. WHAT WILL BE $P(x|\lambda)P(\lambda)$?

$$\ln(P(\lambda)) = \ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\frac{\beta}{\lambda}}\right) = \ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) - (\alpha+1)\ln(\lambda)$$

$$-\frac{\beta}{\lambda} \cancel{\ln}(e)$$

$$\frac{\partial \ln(P(\lambda))}{\partial \lambda} = 0 - \frac{\alpha+1}{\lambda} + \frac{\beta}{\lambda^2}$$

$$P(x|\lambda)P(\lambda) \Rightarrow \ln(P(x|\lambda)P(\lambda)) = \ln(P(x|\lambda)) + \ln(P(\lambda))$$

$$-\frac{N}{\lambda} + \frac{\sum_{n=1}^{N} x_n^2}{\lambda^2} - \frac{\alpha+1}{\lambda} + \frac{\beta}{\lambda^2} = 0 \qquad \frac{(-N-\alpha+1)\lambda + \sum_{n=1}^{N} x_n^2 + \beta}{\lambda^2} = 0$$

$$\lambda = \frac{\sum_{n=1}^{N} x_n^2 + \beta}{N + \alpha - 1}$$

2a) SUPPOSE TO HAVE A SET OF 1-DIMENTIONAL POINTS $x_i \in \{0,1\}$

$D=(x_1,\dots,x_n)$ THAT ARE UNDER THE PROBABILITY DISTRIBUTION

$P(x,q)=\prod_{d=1}^{D} q_D^{x_d}(1-q_D)^{1-x_d}$. WHAT WILL BE THE M.L.E.(q)?

$$\tilde{L}=\ln\left(\prod_{n=1}^{N}\prod_{d=1}^{D} q_D^{x_{nd}}(1-q_D)^{1-x_{nd}}\right)=\sum_{n=1}^{N}\sum_{d=1}^{D}\left(\ln(q_D^{x_{nd}})+\ln(1-q_D)^{1-x_{nd}}\right)$$

$$=\sum_{n=1}^{N}\sum_{d=1}^{D}\left(x_{nd}\ln(q_D)+(1-x_{nd})\ln(1-q_D)\right)$$

$$\frac{\partial\tilde{L}}{\partial q_d}=\sum_{n=1}^{N}\left(\frac{x_{nd}}{q_D}+\frac{1-x_{nd}}{1-q_D}\cdot(-1)\right)=0 \qquad \frac{\sum_{n=1}^{N}x_{nd}}{q_D}-\frac{N-\sum_{n=1}^{N}x_{nd}}{1-q_D}=0$$

$$\frac{\sum_{n=1}^{N}x_{nd}-\cancel{q_D\sum_{n=1}^{N}x_{nd}}-Nq_D+\cancel{q_d\sum_{n=1}^{N}x_{nd}}}{\cancel{q_d(1-q_d)}}=0 \qquad q_D=\frac{\sum_{n=1}^{N}x_{nd}}{N}$$

2b) ASSUME, NOW, THAT $P(q_D)=\frac{1}{\beta(\alpha,\beta)}q_D^{\alpha-1}(1-q_D)^{\beta-1}$.

WHAT WILL BE $P(x|q_D)P(q_D)$?

$$\ln(P(q_D))=\ln\left(\frac{1}{\beta(\alpha+\beta)}+(\alpha-1)\ln(q_D)+(\beta-1)\ln(1-q_D)\right)$$

$$\frac{\partial\ln(P(q_D))}{\partial q_D}=0+\frac{\alpha-1}{q_D}+\frac{\beta-1}{1-q_D}$$

$$P(x|q_D)P(q_D)\Rightarrow\ln(P(x|q_D)P(q_D))=\ln(P(x|q_D))+\ln(q_D)=$$

$$=\frac{\sum_{n=1}^{N}x_{nd}}{q_D}-\frac{N-\sum_{n=1}^{N}x_{nd}}{1-q_D}+\frac{\alpha-1}{q_D}+\frac{\beta-1}{1-q_D}=0$$

$$\frac{\sum_{n=1}^{N} x_{nd}}{q_D} - \frac{N - \sum_{n=1}^{N} x_{nd}}{1-q_D} + \frac{\alpha-1}{q_D} + \frac{\beta-1}{1-q_D}$$

$$\frac{\sum_{n=1}^{N} x_{nd} - N q_D + \alpha - q_D\alpha - 1 + q_D - \beta q_D + q_D}{q_D \cancel{(1-q_D)}} = 0$$

$$q_D = \frac{\sum_{n=1}^{N} x_{nd} + \alpha - 1}{N + \alpha + \beta - 2}$$

EXTRA THEORETICAL CONCEPTS

$$\ln(P(X|\theta)) = \int_z q(z) \ln(X|\theta)\, dz = \int_z q(z) \ln\left(\frac{P(X,Z|\theta)}{P(Z|X,\theta)}\right) dz =$$

$$= \int_z q(z) \ln\left(\frac{P(X,Z|\theta)}{q(z)}\right) dz + \int_z q(z) \ln\left(\frac{q(z)}{P(Z|X,\theta)}\right) dz =$$

DEFINING $KL(P(x)||q(x)) = \int_x P(x) \ln\left(\frac{P(x)}{q(x)}\right) dx \geq 0$ AND THE

ELBO (EVIDENCE LOWER BOUND) $\mathcal{L}(q,\theta) = \int_z q(z) \ln\left(\frac{P(X,Z|\theta)}{q(z)}\right) dz$

$$= \mathcal{L}(q,\theta) + KL(q(z)||P(Z|X,\theta)) \geq \mathcal{L}(q,\theta)$$

E-STEP

$$q_{NEW} = \underset{q}{\operatorname{argmax}}\, \mathcal{L}(q,\theta) = \underset{q}{\operatorname{argmin}}\, KL(q(z)||P(Z|X,\theta)) = P(Z|X,\theta)$$

M-STEP

$$\theta_{NEW} = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(q,\theta) = \underset{\theta}{\operatorname{argmax}} \int_z q(z) \ln\left(\frac{P(X,Z|\theta)}{q(z)}\right) dz$$

$$= \underset{\theta}{\operatorname{argmax}} \int_z q(z) \ln(P(X,Z|\theta))\, dz = \underset{\theta}{\operatorname{argmax}}\, E[\ln(P(X,Z|\theta))]$$

$$P(X_n) = \sum_{i=1}^{K} w_i \mathcal{N}(X_n|\mu_i,\sigma_i)$$

REPEAT UNTIL SATISFIED

E-STEP

$$q(z) = P(Z|X,\theta) = \frac{w_k \mathcal{N}_k(X)}{\sum_{i=1}^{K} w_i \mathcal{N}_i(x)}$$

M-STEP

$$P(x_n, z_n \mid \theta) = \prod_{i=1}^{K} (w_k \mathcal{N}_i(x_n))^{z_{ni}}$$

$$\Rightarrow P(X, Z \mid \theta) = \prod_{n=1}^{N} \prod_{i=1}^{K} (w_i \mathcal{N}_i(x_n))^{z_{ni}}$$

γ_{ni} ARE COMPONENTS OF $P(z_n \mid x_n, \theta)$

$$\ln(P(X, Z \mid \theta)) = \sum_{n=1}^{N} \sum_{i=1}^{K} z_{ni} \ln(w_i \mathcal{N}_i(x_i))$$

$$E_{Z \sim q(Z)}[\ln(P(X, Z \mid \theta))] = \sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{ni} \ln(w_i \mathcal{N}_i(x_i))$$

WE WILL FIND THE OPTIMAL VALUES FOR $\hat{w}_k, \hat{\mathcal{N}}_k, \hat{\sigma}_k^2$

- $$\frac{\partial}{\partial w_i}\left(\sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{ni} \ln(w_i \mathcal{N}_i(x_i)) + \lambda\left(1 - \sum_{i=1}^{K} w_i\right)\right) = 0$$ LAGRANGE MULTIPLIERS

$$\sum_{n=1}^{N} \frac{\gamma_{ni}}{w_i} = \lambda \Rightarrow w_i = \frac{1}{\lambda} \sum_{n=1}^{N} \gamma_{ni}$$

$$\sum_{i=1}^{K} \frac{1}{\lambda} \sum_{n=1}^{N} \gamma_{ni} = 1 \Rightarrow \lambda = N \Rightarrow w_k = \frac{\sum_{n=1}^{N} \gamma_{nk}}{N}$$

- $$\frac{\partial}{\partial \mathcal{N}_k}\left(\sum_{n=1}^{N} \sum_{i=1}^{K} \gamma_{ni} \ln(w_i \mathcal{N}_i(x_i))\right) = 0$$

$$\mathcal{N}(x \mid \mathcal{N}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mathcal{N})^2}{2\sigma^2}} \Rightarrow \ln(\mathcal{N}) = -\ln(\sqrt{2\pi\sigma^2}) - \frac{(x-\mathcal{N})^2}{2\sigma^2}$$

$$\sum_{n=1}^{N} \gamma_{nk} \frac{\partial}{\partial \mathcal{N}_k}\left(-\frac{(x_n - \mathcal{N}_k)^2}{2\sigma_k^2}\right) = 0$$

$$\sum_{n=1}^{N} \gamma_{nk}\left(\frac{x_n - \mathcal{N}_k}{\sigma_k^2}\right) = 0 \Rightarrow \sum_{n=1}^{N} \gamma_{nk}(x_n - \mathcal{N}_k) = 0$$

$$\mathcal{N}_k = \frac{\sum_{n=1}^{N} \gamma_{nk} x_n}{\sum_{n=1}^{N} \gamma_{nk}}$$

$$\frac{\partial}{\partial \sigma_k^2}\left(\sum_{n=1}^{N}\sum_{i=1}^{K}\gamma_{ni}\ln\left(w_i N_i\left(x_i^{(n)}\right)\right)\right)=0$$

$$\ln(N) = -\ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\sum_{n=1}^{N}\gamma_{nk}\frac{\partial}{\partial \sigma_k^2}\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x^{(n)}-\mu)^2}{2\sigma^2}\right)$$

$$\sum_{n=1}^{N}\gamma_{nk}\left(-\frac{1\cdot\sigma^2}{\cancel{4}\pi\cancel{\sigma}\cancel{2}}\cdot(2\cancel{\pi}) + \frac{(x^{(n)}-\mu)^2}{\cancel{4}\cancel{\sigma^3}}\right)=0$$

$$\sum_{n=1}^{N}\gamma_{nk}\left(-\frac{2}{\pi}\sigma^2 + (x^{(n)}-\mu)^2\right)=0$$

$$\hat{\sigma}^2 = \frac{\sum_{n=1}^{N}\gamma_{nk}(x^{(n)}-\mu)(x^{(n)}-\mu)^T}{\sum_{n=1}^{N}\gamma_{nk}}$$