GEMMULEM Gaussian Mixture Deconvolution Description and Illustrative Example Derivation

Micah Thornton & Chanhee Park (2022)
Lyda Hill Department of Bioinformatics

General Gaussian Mixture Example.

- Suppose we use a spectrophotometer to measure a continuous index of reflectivity, r_i ($r_i \in \mathbb{R}$), of the i^{th} (of N total) marble.
  - For each marble ($\forall\ i,\ i \in \mathbb{Z}$), there exists a true reflectivity, r_i, (we are ignoring measurement error in this simplifying example) and true color, x_i.

- Let x_i be the true color of the i^{th} marble.
- Let r_i be the observed reflectivity of the i^{th} marble.
- Assume the reflectivity of marbles are modeled as a Gaussian mixture, with parameters depending on the true color of each marble.

$$P(R_i = r_i \mid X_i = x_i) = \begin{Bmatrix} \frac{1}{\sigma_R\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_R}{2\sigma_R}\right\}^2}, & x_i \equiv R \\ \frac{1}{\sigma_G\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_G}{2\sigma_G}\right\}^2}, & x_i \equiv G \\ \frac{1}{\sigma_B\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_B}{2\sigma_B}\right\}^2}, & x_i \equiv B \end{Bmatrix}$$

$$P(X_i = x_i) = \begin{Bmatrix} \gamma_R, & x_i \equiv R \\ \gamma_G, & x_i \equiv G \\ \gamma_B, & x_i \equiv B \end{Bmatrix}$$

- We want to learn about the parameters (mean, variance, and probabilities) associated with the above model.

$$\boldsymbol{\theta} = \begin{bmatrix} \overrightarrow{\boldsymbol{\theta_R}} \\ \overrightarrow{\boldsymbol{\theta_G}} \\ \overrightarrow{\boldsymbol{\theta_B}} \end{bmatrix}, \qquad \overrightarrow{\boldsymbol{\theta_k}} = \begin{Bmatrix} \gamma_k \\ \mu_k \\ \sigma_k \end{Bmatrix}, \qquad k \in \{`R', `B', `G'\}$$

- But we observe only the r_i, and do not know x_i, hence we are working with the coarsened distribution on R_i.

$$R_i = r_i \in \mathbb{R} \iff \mathfrak{S}_{R_i} \equiv \mathbb{R}$$

$$P(R_i = r_i) = \frac{\gamma_R}{\sigma_R\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_R}{2\sigma_R}\right\}^2} + \frac{\gamma_G}{\sigma_G\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_G}{2\sigma_G}\right\}^2} + \frac{\gamma_B}{\sigma_B\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_B}{2\sigma_B}\right\}^2}$$

- To make inference on the parameters of the distributional model assumed here, we would like to know $\overrightarrow{\boldsymbol{T_k}}$, the sum of the reflectivity and square of reflectivity of all marbles of color $X_i = x_i$.

$$\overrightarrow{\boldsymbol{T_k}} = \begin{bmatrix} \sum_{i=1}^{N} \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i \cdot \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i^2 \cdot \mathbb{1}(X_i = k) \end{bmatrix}$$

- We can inspect the **full expectation** of this quantity under **the complete data model** (R_i and X_i known $\forall\ i \in [0, N]$).

$$E\left[\overrightarrow{\boldsymbol{T_k}} \mid (\vec{\boldsymbol{R}}, \vec{\boldsymbol{X}})\right] = \boldsymbol{E}\begin{bmatrix} \sum_{i=1}^{N} \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i \cdot \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i^2 \cdot \mathbb{1}(X_i = k) \end{bmatrix} = \begin{bmatrix} N \cdot E[\mathbb{1}(X_i = k)] \\ N \cdot E[R_i | X_i = k] \cdot E[\mathbb{1}(X_i = k)] \\ N \cdot E\left[R_i^2 \middle| X_i = x_i\right] \cdot E[\mathbb{1}(X_i = k)] \end{bmatrix} = \begin{bmatrix} N \cdot \gamma_k \\ N \cdot r_i \cdot \frac{1}{\sigma_k\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_k}{2\sigma_k}\right\}^2} \cdot \gamma_k \\ N \cdot r_i^2 \cdot \frac{1}{\sigma_k\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_k}{2\sigma_k}\right\}^2} \cdot \gamma_k \end{bmatrix}$$

- Unfortunately, as is usually the case, we do not have complete information, but instead coarsened observations, and in most cases completely coarsened (ie. r_i is observed, but It is only known that $x_i \in \mathfrak{S}_{X_i} \equiv \{R, G, B\}$) we can therefore inspect the **conditional expectation** of $\overrightarrow{T_k}$ **under the coarsened data model** (only R_i known)

$$E^{(t)}\left[\overrightarrow{\boldsymbol{T_k}} \mid \vec{\boldsymbol{R}}\right] = \boldsymbol{E}\begin{bmatrix} \sum_{i=1}^{N} \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i \cdot \mathbb{1}(X_i = k) \\ \sum_{i=1}^{N} R_i^2 \cdot \mathbb{1}(X_i = k) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} P_{\theta^{(t-1)}}(X_i = k \mid R_i = r_i) \\ \sum_{i=1}^{N} r_i \cdot P_{\boldsymbol{\theta}^{(t-1)}}(R_i = r_i) \\ \sum_{i=1}^{N} r_i^2 \cdot P_{\boldsymbol{\theta}^{(t-1)}}(R_i = r_i) \end{bmatrix} = \begin{bmatrix} N \cdot \widehat{\gamma_k}^{(t)} \\ N \cdot \widehat{\mu_k}^{(t)} \\ N \cdot \left(\widehat{\sigma_k^2}^{(t)} + \widehat{\mu_k^2}^{(t)}\right) \end{bmatrix} = \begin{bmatrix} T_{k(1)}^{(t)} \\ T_{k(2)}^{(t)} \\ T_{k(3)}^{(t)} \end{bmatrix} \qquad [E - Step]$$

- Note that,

$$P_{\theta^{(t-1)}}(X_i = k \mid R_i = r_i) = P_{\theta^{(t-1)}}(R_i = r_i \mid X_i = k) \cdot \frac{P_{\theta^{(t-1)}}(X_i = k)}{P_{\theta^{(t-1)}}(R_i = r_i)}$$

- Can be computed directly.

- Given an initial guess (or the previous iteration result) for the parameters (denoted $\boldsymbol{\theta}^{(\boldsymbol{t-1})}$)in the subscript of the probability statement above) the conditional expectation can be computed.
- Finally, the parameters which maximize the likelihood of observing the computed statistics may be computed.

$$\widehat{\gamma_k^{(t)}} = N^{-1} \cdot T_{k(1)}^{(t)} = \sum_{i=1}^{N} \frac{\frac{1}{\sigma_k^{(t-1)}\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_k^{(t-1)}}{2\sigma_k^{(t-1)}}\right\}^2} \cdot \gamma_k^{(t-1)}}{N \cdot \sum_{l \in \{R,G,B\}} \left(\frac{1}{\sigma_l^{(t-1)}\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_l^{(t-1)}}{2\sigma_l^{(t-1)}}\right\}^2} \cdot \gamma_l^{(t-1)}\right)}$$

$$\widehat{\mu_k^{(t)}} = N^{-1} \cdot T_{k(2)}^{(t)} = N^{-1} \cdot \sum_{i=1}^{N} r_i \cdot \sum_{l \in \{R,G,B\}} \left(\frac{1}{\sigma_l^{(t-1)}\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_l^{(t-1)}}{2\sigma_l^{(t-1)}}\right\}^2} \cdot \gamma_l^{(t-1)}\right) \qquad [M - Step]$$

$$\widehat{\sigma^2}_k^{(t)} = \left\{N^{-1} \cdot \sum_{i=1}^{N} r_i^2 \cdot \sum_{l \in \{R,G,B\}} \left(\frac{1}{\sigma_l^{(t-1)}\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_l^{(t-1)}}{2\sigma_l^{(t-1)}}\right\}^2} \cdot \gamma_l^{(t-1)}\right)\right\} - \left\{N^{-1} \cdot \sum_{i=1}^{N} r_i \cdot \sum_{l \in \{R,G,B\}} \left(\frac{1}{\sigma_l^{(t-1)}\sqrt{2\pi}} \cdot e^{\left\{-\frac{r_i-\mu_l^{(t-1)}}{2\sigma_l^{(t-1)}}\right\}^2} \cdot \gamma_l^{(t-1)}\right)\right\}^2$$

Notes about the GMM deconvolution implementation in GEMMULEM.

- The above procedure is known as density deconvolution and in this case it is applied to a known (or pre-supplied) number of Gaussians.
- The procedure can be generalized for any density function, and even for arbitrary densities (mixtures of mixtures).
- This illustrating example shows how the EM steps can be derived for a simple univariate Gaussian Mixture model, but this algorithm is already implemented and available to you for general use in GEMMULEM.
- Typically the E-step and the M-step as shown above will alternate producing and iteratively refined estimate of the parameters of the model from which these data come.