

## Estimation Using EM

### 1. Model of BGG

BGG assumes that the observed signals of all the probes in either dye ( $y_{ijk_j}$ ) follow Gamma distribution, while the scale parameter ( $\theta$ ) of this Gamma distribution also follows a Gamma distribution. Then, a global probability ( $p$ ) indicating the proportion of differential probes and non-differential probes between two samples is defined to integrate Bernoulli distributed properties over the Gamma-Gamma distributions. ( $i$  is the index of each probe ( $n$  is maximum of  $i$ );  $j$  is the label of either sample ( $r$  is red dye and  $g$  is green dye);  $k_j$  represents the index of each replicate in  $j$  sample;  $K_j$  equals  $\sum k_j$ )

The null hypothesis  $H_0$  is that the probe doesn't show differential between samples. The marginal probability  $p_0(y_{ir}, y_{ig})$  is derived as following.

$$H_0: \theta_{ir} = \theta_{ig} = \theta_i \quad (r \neq g)$$

$$y_{irk_r} \sim \Gamma(a, \theta_i), y_{igk_g} \sim \Gamma(a, \theta_i), \theta_i \sim \Gamma(a_0, \nu)$$

$$p_0(y_{ir}, y_{ig}) = p(y_{irk_r}, y_{igk_g}, \theta_i | a, a_0, \nu) = \frac{\Gamma(aK_r + aK_g + a_0)}{\Gamma(a_0)\Gamma^{K_r+K_g}(a)} \cdot \frac{\nu^{a_0} (\prod_{k_r} y_{irk_r} \prod_{k_g} y_{igk_g})^{a-1}}{(\nu + \sum_{k_r} y_{irk_r} + \sum_{k_g} y_{igk_g})^{aK_r+aK_g+a_0}}$$

The alternative hypothesis  $H_A$  is that the probe differentially regulated between samples. The marginal probability  $p_A(y_{ir}, y_{ig})$  is derived as following.

$$H_A: \theta_{ir} \neq \theta_{ig} \quad (r \neq g)$$

$$y_{irk_r} \sim \Gamma(a, \theta_{ir}), y_{igk_g} \sim \Gamma(a, \theta_{ig}), \theta_{ir} \sim \Gamma(a_0, \nu), \theta_{ig} \sim \Gamma(a_0, \nu)$$

$$p_A(y_{ir}, y_{ig}) = p(y_{irk_r}, y_{igk_g}, \theta_{ir}, \theta_{ig} | a, a_0, \nu) = \frac{\Gamma(aK_r+a_0)\Gamma(aK_g+a_0)}{\Gamma^2(a_0)\Gamma^{K_r+K_g}(a)} \cdot \frac{\nu^{2a_0} (\prod_{k_r} y_{irk_r} \prod_{k_g} y_{igk_g})^{a-1}}{(\nu + \sum_{k_r} y_{irk_r})^{aK_r+a_0} (\nu + \sum_{k_g} y_{igk_g})^{aK_g+a_0}}$$

When setting  $z_i$  as the probability of differentially regulated for  $i^{th}$  probe and  $p$  as the success probability of Bernoulli distribution, the log-likelihood is

$$l_c(a, a_0, \nu, p) = \sum_i \{z_i \log p_A(y_{ir}, y_{ig}) + (1 - z_i) \log p_0(y_{ir}, y_{ig}) + z_i \log p + (1 - z_i) \log(1 - p)\}$$

As the summary of this mixture model, the observed values, parameters and missing values are  $\{y_{irk_r}, y_{igk_g}\}$ ,  $\{a, a_0, \nu, p\}$  and  $\{z_i\}$ , respectively.

**E-Step:**

$$\hat{z}_i = P(z_i = 1 | y_{irk_r}, y_{igk_g}, a, a_0, \nu, p) = \frac{p \cdot p_A(y_{ir}, y_{ig})}{p \cdot p_A(y_{ir}, y_{ig}) + (1-p) p_0(y_{ir}, y_{ig})}$$

**M-step:**

$$\hat{p} = \frac{2 + \sum_i \hat{z}_i}{2 \cdot 2 + n}$$

$$(\hat{a}, \hat{a}_0, \hat{\nu}) = \operatorname{argmax}_{a, a_0, \nu} l_c(a, a_0, \nu, \hat{p})$$

Parameters  $\hat{a}, \hat{a}_0, \hat{\nu}$  are numerically optimized by function **nlminb** in R.

EM iterations terminated when reaching the convergence which was defined as no more than 0.01% changes on log-likelihood.

### 2. Model of BNGG

On the basis of BGG, BNKG further considered the normal distributed biases between different probes ( $\delta_{ijk_j}$ ), with mean as 0 and standard deviation as  $\tau_i$ . The model we used here is given:

$$y_{ijk_j} = \eta_{ij} + \delta_{ijk_j}, \delta_{ijk_j} \sim N(0, \tau_i^2)$$

The null hypothesis  $H_0$  is that the probe doesn't show differential between samples. The marginal probability  $p_0(y_{ir}, y_{ig})$  is derived as following.

$$\begin{aligned} H_0: \theta_{ir} = \theta_{ig} = \theta_i \quad (r \neq g) \\ y_{irk_r} \sim N(\eta_{ir}, \tau_i^2), y_{igk_g} \sim N(\eta_{ig}, \tau_i^2), \eta_{ir} \sim \Gamma(a, \theta_i), \eta_{ig} \sim \Gamma(a, \theta_i), \theta_i \sim \Gamma(a_0, \nu) \\ p_0(y_{ir}, y_{ig}) = \frac{\Gamma(2a + a_0)}{\Gamma^2(a)\Gamma(a_0)} \cdot \nu^{a_0} \cdot \Delta_{ir} \cdot \Delta_{ig} \cdot \mathbf{E} \left[ \frac{\eta_{ir}^{a-1} \eta_{ig}^{a-1}}{(\eta_{ir} + \eta_{ig} + \nu)^{2a+a_0}} \right]_{\text{TN} \left( \frac{\sum_{k_r} \{y_{irk_r}\}}{K_r}, \frac{\tau_i^2}{K_r}, \frac{\sum_{k_g} \{y_{igk_g}\}}{K_g}, \frac{\tau_i^2}{K_g} | \eta_{ir}, \eta_{ig} > 0 \right)} \end{aligned}$$

Where **TN** stands for 'Truncated Normal Distribution'; **E** stands for 'Expectation';  $\Delta_{ij}$  is given:

$$\Delta_{ij} = \sqrt{\frac{1}{2^{K_j-1} \pi^{K_j-1} \tau_i^{2(K_j-1)} K_j}} * e^{\frac{(\sum_{k_j} \{y_{ijk_j}\})^2 - K_j \sum_{k_j} \{y_{ijk_j}^2\}}{2K_j \tau_i^2}}$$

The alternative hypothesis  $H_A$  is that the probe differentially regulated between samples. The marginal probability  $p_A(y_{ir}, y_{ig})$  is derived as following.

$$\begin{aligned} H_A: \theta_{ir} \neq \theta_{ig} \quad (r \neq g) \\ y_{irk_r} \sim N(\eta_{ir}, \tau_i^2), y_{igk_g} \sim N(\eta_{ig}, \tau_i^2), \eta_{ir} \sim \Gamma(a, \theta_{ir}), \eta_{ig} \sim \Gamma(a, \theta_{ig}), \theta_{ir} \sim \Gamma(a_0, \nu), \theta_{ig} \sim \Gamma(a_0, \nu) \\ p_A(y_{ir}, y_{ig}) = \left\{ \frac{\Gamma(a+a_0)}{\Gamma(a)\Gamma(a_0)} \right\}^2 \cdot \nu^{2a_0} \cdot \Delta_{ir} \cdot \Delta_{ig} \cdot \mathbf{E} \left[ \frac{(\eta_{ir})^{a-1}}{(\eta_{ir} + \nu)^{a+a_0}} \right]_{\text{TN} \left( \frac{\sum_{k_r} \{y_{irk_r}\}}{K_r}, \frac{\tau_i^2}{K_r} | \eta_{ir} > 0 \right)} \mathbf{E} \left[ \frac{(\eta_{ig})^{a-1}}{(\eta_{ig} + \nu)^{a+a_0}} \right]_{\text{TN} \left( \frac{\sum_{k_g} \{y_{igk_g}\}}{K_g}, \frac{\tau_i^2}{K_g} | \eta_{ig} > 0 \right)} \end{aligned}$$

When setting  $z_i$  as the probability of differentially regulated for  $i^{th}$  probe and  $p$  as the success probability of Bernoulli distribution, the log-likelihood is

$$\mathcal{L}_c(\tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p) = \sum_i \{z_i \log p_A(y_{ir}, y_{ig}) + (1 - z_i) \log p_0(y_{ir}, y_{ig}) + z_i \log p + (1 - z_i) \log(1 - p)\}$$

As the summary of this mixture model, the observed values, parameters and missing values are  $\{y_{irk_r}, y_{igk_g}\}$ ,  $\{\tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p\}$  and  $\{z_i\}$ , respectively.

**E-Step:**

$$\hat{z}_i = P(z_i = 1 | y_{irk_r}, y_{igk_g}, \tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p) = \frac{p \cdot p_A(y_{ir}, y_{ig})}{p \cdot p_A(y_{ir}, y_{ig}) + (1-p) p_0(y_{ir}, y_{ig})}$$

**M-step:**

$$\begin{aligned} \hat{p} &= \frac{2 + \sum_i \hat{z}_i}{2 \cdot 2 + n} \\ (\hat{\tau}_1^2, \dots, \hat{\tau}_n^2, \hat{a}, \hat{a}_0, \hat{\nu}) &= \arg \max_{\tau_1^2, \dots, \tau_n^2, a, a_0, \nu} \mathcal{L}_c(\tau_1^2, \dots, \tau_n^2, a, a_0, \nu, \hat{p}) \end{aligned}$$

Parameters  $\hat{\tau}_1^2, \dots, \hat{\tau}_n^2, \hat{a}, \hat{a}_0, \hat{\nu}$  are numerically optimized by function **nlminb** in R. In detail, Hill Climbing was used to optimize these parameters for the iteration of each M-step, to efficiently cut down the time cost of function **nlminb**.

EM iterations terminated when reaching the convergence which was defined as no more than 0.01% changes on log-likelihood.

### 3. Model of BNNGG

In addition to BNGG, pixel biases ( $\varepsilon_{ijk_j}$ ) of probes were further considered in model of BNNGG. We assumed that it follows normal distribution with mean as 0 and standard deviation as  $\sigma_{ijk_j}$ . The model we used here is given:

$$y_{ijk_j} = \eta_{ij} + \varepsilon_{ijk_j} + \delta_{ijk_j}, \varepsilon_{ijk_j} \sim N(0, \sigma_{ijk_j}^2), \delta_{ijk_j} \sim N(0, \tau_i^2)$$

The null hypothesis  $H_0$  is that the probe doesn't show differential between samples. The marginal probability  $p_0(y_{ir}, y_{ig})$  is derived as following.

$$\begin{aligned} H_0: \theta_{ir} &= \theta_{ig} = \theta_i \quad (r \neq g) \\ y_{irk_r} &\sim N(\eta_{ir}, \sigma_{irk_r}^2 + \tau_i^2), y_{igk_g} \sim N(\eta_{ig}, \sigma_{igk_g}^2 + \tau_i^2), \eta_{ir} \sim \Gamma(a, \theta_i), \eta_{ig} \sim \Gamma(a, \theta_i), \theta_i \sim \Gamma(a_0, \nu) \\ p_0(y_{ir}, y_{ig}) &= \frac{\Gamma(2a+a_0)}{\Gamma^2(a)\Gamma(a_0)} \cdot \nu^{a_0} \cdot \Delta_{ir} \cdot \Delta_{ig} \cdot \\ E \left[ \frac{(\eta_{ir}\eta_{ig})^{a-1}}{(\eta_{ir}+\eta_{ig}+\nu)^{2a+a_0}} \right] &= \text{TN} \left( \frac{\sum_{k_r} \{y_{irk_r} \prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}} \cdot \frac{\prod_{k_r} (\sigma_{irk_r}^2 + \tau_i^2)}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}} \cdot \frac{\sum_{k_g} \{y_{igk_g} \prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}} \cdot \frac{\prod_{k_g} (\sigma_{igk_g}^2 + \tau_i^2)}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}} | \eta_{ir}, \eta_{ig} > 0 \right) \end{aligned}$$

Where  $\Delta_{ij}$  is given:

$$\Delta_{ij} = \sqrt{\frac{1}{2^{K_j-1} \pi^{K_j-1} \sum_{k_j} \{\prod_{m \neq k_j} (\sigma_{ijm}^2 + \tau_i^2)\}}} * e^{-\frac{\sum_{l_j \neq k_j, l_j < k_j} (y_{ijl_j} - y_{ijk_j})^2 \prod_{m \neq l_j, m \neq k_j} (\sigma_{ijm}^2 + \tau_i^2)}{2 * \sum_{k_j} \{\prod_{m \neq k_j} (\sigma_{ijm}^2 + \tau_i^2)\}}}$$

The alternative hypothesis  $H_A$  is that the probe differentially regulated between samples. The marginal probability  $p_A(y_{ir}, y_{ig})$  is derived as following.

$$\begin{aligned} H_A: \theta_{ir} &\neq \theta_{ig} \quad (r \neq g) \\ y_{irk_r} &\sim N(\eta_{ir}, \sigma_{irk_r}^2 + \tau_i^2), y_{igk_g} \sim N(\eta_{ig}, \sigma_{igk_g}^2 + \tau_i^2), \eta_{ir} \sim \Gamma(a, \theta_{ir}), \eta_{ig} \sim \Gamma(a, \theta_{ig}), \theta_{ir} \sim \Gamma(a_0, \nu), \theta_{ig} \sim \Gamma(a_0, \nu) \\ p_A(y_{ir}, y_{ig}) &= \frac{\Gamma(a+a_0)}{\Gamma(a)\Gamma(a_0)}^2 \cdot \nu^{2a_0} \cdot \Delta_{ir} \cdot \Delta_{ig} \cdot \\ E \left[ \frac{(\eta_{ir})^{a-1}}{(\eta_{ir}+\nu)^{a+a_0}} \right] &= \text{TN} \left( \frac{\sum_{k_r} \{y_{irk_r} \prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}} \cdot \frac{\prod_{k_r} (\sigma_{irk_r}^2 + \tau_i^2)}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma_{irm}^2 + \tau_i^2)\}} | \eta_{ir} > 0 \right) \\ E \left[ \frac{(\eta_{ig})^{a-1}}{(\eta_{ig}+\nu)^{a+a_0}} \right] &= \text{TN} \left( \frac{\sum_{k_g} \{y_{igk_g} \prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}} \cdot \frac{\prod_{k_g} (\sigma_{igk_g}^2 + \tau_i^2)}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma_{igm}^2 + \tau_i^2)\}} | \eta_{ig} > 0 \right) \end{aligned}$$

When setting  $z_i$  as the probability of differentially regulated for  $i^{th}$  probe and  $p$  as the success probability of Bernoulli distribution, the log-likelihood is

$$l_c(\tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p) = \sum_i \{z_i \log p_A(y_{ir}, y_{ig}) + (1 - z_i) \log p_0(y_{ir}, y_{ig}) + z_i \log p + (1 - z_i) \log(1 - p)\}$$

As the summary of this mixture model, the observed values, parameters and missing values are  $\{y_{irk_r}, y_{igk_g}, \sigma_{irk_r}^2, \sigma_{igk_g}^2\}$ ,  $\{\tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p\}$  and  $\{z_i\}$ , respectively.

**E-Step:**

$$\hat{z}_i = P(z_i = 1 | y_{irk_r}, y_{igk_g}, \sigma_{irk_r}^2, \sigma_{igk_g}^2, \tau_1^2, \dots, \tau_n^2, a, a_0, \nu, p) = \frac{p \cdot p_A(y_{ir}, y_{ig})}{p \cdot p_A(y_{ir}, y_{ig}) + (1-p) p_0(y_{ir}, y_{ig})}$$

**M-step:**

$$\hat{p} = \frac{2 + \sum_i \hat{z}_i}{2 \cdot 2 + n}$$

$$(\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \widehat{a}, \widehat{a_0}, \widehat{v}) = \operatorname{argmax}_{\tau_1^2, \dots, \tau_n^2, a, a_0, v} \mathbf{l}_c(\tau_1^2, \dots, \tau_n^2, a, a_0, v, \hat{p})$$

Parameters  $\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \widehat{a}, \widehat{a_0}, \widehat{v}$  are numerically optimized by function **nlminb** in R. In detail, Hill Climbing was used to optimize these parameters for the iteration of each M-step, to efficiently cut down the time cost of function **nlminb**.

EM iterations terminated when reaching the convergence which was defined as no more than 0.01% changes on log-likelihood.

#### 4. Model of BLNN

In model BLNN, all the observed signals ( $y_{ijk_j}$ ) were firstly logarithmically transformed to  $y'_{ijk_j}$ . We assumed that  $y'_{ijk_j}$  consists of the true intensity of each probe ( $\eta_{ij}$ ) and normal distributed biases between different probes ( $\delta_{ijk_j}$ ). The model we used here is given:

$$y'_{ijk_j} = \eta_{ij} + \delta_{ijk_j}, \delta_{ijk_j} \sim N(0, \tau_i^2)$$

The null hypothesis  $H_0$  is that the probe doesn't show differential between samples. The marginal probability  $p_0(y'_{ir}, y'_{ig})$  is derived as following.

$$\begin{aligned} H_0: \eta_{ir} &= \eta_{ig} = \eta_i \quad (r \neq g) \\ y'_{irk_r} &\sim N(\eta_i, \tau_i^2), y'_{igk_g} \sim N(\eta_i, \tau_i^2), \eta_i \sim N(\mu, \varphi^2) \\ p_0(y'_{ir}, y'_{ig}) &= \frac{\Delta_{ir} \Delta_{ig}}{2\pi\sqrt{AB + BC + AC}} * e^{-\frac{A(E-F)^2 + B(D-F)^2 + C(D-E)^2}{2(AB + BC + AC)}} \end{aligned}$$

where  $A, B, C, D, E, F$  and  $\Delta_{ij}$  are given:

$$\begin{aligned} A &= \frac{\tau_i^2}{K_r}; B = \frac{\tau_i^2}{K_g}; C = \varphi^2; D = \frac{\sum_{k_r} \{y'_{irk_r}\}}{K_r}; E = \frac{\sum_{k_g} \{y'_{igk_g}\}}{K_g}; F = \mu \\ \Delta_{ij} &= \sqrt{\frac{1}{2^{K_j-1} \pi^{K_j-1} \tau_i^{2(K_j-1)} K_j}} * e^{-\frac{(\sum_{k_j} \{y'_{ijk_j}\})^2 - K_j \sum_{k_j} \{y'_{ijk_j}\}^2}{2K_j \tau_i^2}} \end{aligned}$$

The alternative hypothesis  $H_A$  is that the probe differentially regulated between samples. The marginal probability  $p_A(y'_{ir}, y'_{ig})$  is derived as following.

$$\begin{aligned} H_A: \eta_{ir} &\neq \eta_{ig} \quad (r \neq g) \\ y'_{irk_r} &\sim N(\eta_{ir}, \tau_i^2), y'_{igk_g} \sim N(\eta_{ig}, \tau_i^2), \eta_{ir} \sim N(\mu, \varphi^2), \eta_{ig} \sim N(\mu, \varphi^2) \\ p_A(y'_{ir}, y'_{ig}) &= \frac{\Delta_{ir} \Delta_{ig}}{2\pi\sqrt{(A+C)(B+C)}} e^{-\left(\frac{(D-F)^2}{2(A+C)} + \frac{(E-F)^2}{2(B+C)}\right)} \end{aligned}$$

When setting  $z_i$  as the probability of differentially regulated for  $i^{th}$  probe and  $p$  as the success probability of Bernoulli distribution, the log-likelihood is

$$\mathbf{l}_c(\tau_1^2, \dots, \tau_i^2, \dots, \tau_n^2, \mu, \varphi^2, p) = \sum_i \{z_i \log p_A(y'_{ir}, y'_{ig}) + (1 - z_i) \log p_0(y'_{ir}, y'_{ig}) + z_i \log p + (1 - z_i) \log(1 - p)\}$$

As the summary of this mixture model, the observed values, parameters and missing values are  $\{y'_{irk_r}, y'_{igk_g}\}$ ,  $\{\tau_1^2, \dots, \tau_i^2, \dots, \tau_n^2, \mu, \varphi^2, p\}$  and  $\{z_i\}$ , respectively.

**E-Step:**

$$\hat{z}_i = P(z_i = 1 | y'_{irk_r}, y'_{igk_g}, \tau_1^2, \dots, \tau_n^2, \mu, \varphi^2, p) = \frac{p \cdot p_A(y'_{ir}, y'_{ig})}{p \cdot p_A(y'_{ir}, y'_{ig}) + (1-p) p_0(y'_{ir}, y'_{ig})}$$

**M-step:**

$$\hat{p} = \frac{2 + \sum_i \hat{z}_i}{2 \cdot 2 + n}$$

$$(\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \widehat{\mu}, \widehat{\varphi^2}) = \operatorname{argmax}_{\tau_1^2, \dots, \tau_n^2, \mu, \varphi^2} l_c(\tau_1^2, \dots, \tau_n^2, \mu, \varphi^2, \hat{p})$$

Parameters  $\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \widehat{\mu}, \widehat{\varphi^2}$  are numerically optimized by function **nlimb** in R. In detail, Hill Climbing was used to optimize these parameters for the iteration of each M-step, to efficiently cut down the time cost of function **nlimb**.

EM iterations terminated when reaching the convergence which was defined as no more than 0.01% changes on log-likelihood.

## 5. Model of BLNNN

In addition to BLNN, pixel biases ( $\varepsilon_{ijk_j}$ ) of probes with standard deviation  $\sigma_{ijk_j}$  were further considered in model of BLNNN.  $\varepsilon_{ijk_j}$  need to be transferred to coefficient of variations  $\varepsilon'_{ijk_j}$ , as corresponding of logarithmic transformation from  $y_{ijk_j}$  to  $y'_{ijk_j}$ . We assumed that  $\varepsilon'_{ijk_j}$  follow normal distribution with mean as 0 and standard deviation as  $\sigma'_{ijk_j}$ , which are derived from microarray. The model we used here is given:

$$y'_{ijk_j} = \eta_{ij} + \varepsilon'_{ijk_j} + \delta_{ijk_j}, \varepsilon'_{ijk_j} \sim N(0, \sigma'^2_{ijk_j}), \delta_{ijk_j} \sim N(0, \tau_i^2)$$

The null hypothesis  $H_0$  is that the probe doesn't show differential between samples. The marginal probability  $p_0(y'_{ir}, y'_{ig})$  is derived as following.

$$H_0: \eta_{ir} = \eta_{ig} = \eta_i \quad (r \neq g)$$

$$y'_{irk_r} \sim N(\eta_i, \sigma'^2_{irk_r} + \tau_i^2), y'_{igk_g} \sim N(\eta_i, \sigma'^2_{igk_g} + \tau_i^2), \eta_i \sim N(\mu, \varphi^2)$$

$$p_0(y'_{ir}, y'_{ig}) = \frac{\Delta_{ir} \Delta_{ig}}{2\pi \sqrt{AB + BC + AC}} * e^{-\frac{A(E-F)^2 + B(D-F)^2 + C(D-E)^2}{2(AB + BC + AC)}}$$

where  $A, B, C, D, E, F$  and  $\Delta_{ij}$  are given:

$$A = \frac{\prod_{k_r} (\sigma'^2_{irk_r} + \tau_i^2)}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma'^2_{irm} + \tau_i^2)\}}; B = \frac{\prod_{k_g} (\sigma'^2_{igk_g} + \tau_i^2)}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma'^2_{igm} + \tau_i^2)\}}; C = \varphi^2;$$

$$D = \frac{\sum_{k_r} \{y_{irk_r} \prod_{m \neq k_r} (\sigma'^2_{irm} + \tau_i^2)\}}{\sum_{k_r} \{\prod_{m \neq k_r} (\sigma'^2_{irm} + \tau_i^2)\}}; E = \frac{\sum_{k_g} \{y_{igk_g} \prod_{m \neq k_g} (\sigma'^2_{igm} + \tau_i^2)\}}{\sum_{k_g} \{\prod_{m \neq k_g} (\sigma'^2_{igm} + \tau_i^2)\}}; F = \mu$$

$$\Delta_{ij} = \sqrt{\frac{1}{2^{K_j-1} \pi^{K_j-1} \sum_{k_j} \{\prod_{m \neq k_j} (\sigma'^2_{ijm} + \tau_i^2)\}}} * e^{-\frac{\sum_{l_j \neq k_j, l_j < k_j} (y_{ijl_j} - y_{ijk_j})^2 \prod_{m \neq l_j, m \neq k_j} (\sigma'^2_{ijm} + \tau_i^2)}{2 * \sum_{k_j} \{\prod_{m \neq k_j} (\sigma'^2_{ijm} + \tau_i^2)\}}}$$

The alternative hypothesis  $H_A$  is that the probe differentially regulated between samples. The marginal probability  $p_A(y'_{ir}, y'_{ig})$  is derived as following.

$$H_A: \eta_{ir} \neq \eta_{ig} \quad (r \neq g)$$

$$y'_{irk_r} \sim N(\eta_{ir}, \sigma'^2_{irk_r} + \tau_i^2), y'_{igk_g} \sim N(\eta_{ig}, \sigma'^2_{igk_g} + \tau_i^2), \eta_{ir} \sim N(\mu, \varphi^2), \eta_{ig} \sim N(\mu, \varphi^2)$$

$$p_A(y'_{ir}, y'_{ig}) = \frac{\Delta_{ir}\Delta_{ig}}{2\pi\sqrt{(A+C)(B+C)}} e^{-\frac{(D-F)^2}{2(A+C)} + \frac{(E-F)^2}{2(B+C)}}$$

When setting  $z_i$  as the probability of differentially regulated for  $i^{th}$  probe and  $p$  as the success probability of Bernoulli distribution, the log-likelihood is

$$l_c(\tau_1^2, \dots, \tau_i^2, \dots, \tau_n^2, \mu, \varphi^2, p) = \sum_i \{z_i \log p_A(y'_{ir}, y'_{ig}) + (1 - z_i) \log p_0(y'_{ir}, y'_{ig}) + z_i \log p + (1 - z_i) \log(1 - p)\}$$

As the summary of this mixture model, the observed values, parameters and missing values are  $\{y'_{irk_r}, y'_{igk_g}, \sigma'^2_{irk_1}, \sigma'^2_{igk_2}\}$ ,  $\{\tau_1^2, \dots, \tau_i^2, \dots, \tau_n^2, \mu, \varphi^2, p\}$  and  $\{z_i\}$ , respectively.

**E-Step:**

$$\hat{z}_i = P(z_i = 1 | y'_{irk_r}, y'_{igk_g}, \sigma'^2_{irk_1}, \sigma'^2_{igk_2}, \tau_1^2, \dots, \tau_n^2, \mu, \varphi^2, p) = \frac{p \cdot p_A(y'_{ir}, y'_{ig})}{p \cdot p_A(y'_{ir}, y'_{ig}) + (1-p) p_0(y'_{ir}, y'_{ig})}$$

**M-step:**

$$\hat{p} = \frac{2 + \sum_i \hat{z}_i}{2 \cdot 2 + n}$$

$$(\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \hat{\mu}, \widehat{\varphi^2}) = \operatorname{argmax}_{\tau_1^2, \dots, \tau_n^2, \hat{\mu}, \varphi^2} l_c(\tau_1^2, \dots, \tau_n^2, \mu, \varphi^2, \hat{p})$$

Parameters  $\widehat{\tau_1^2}, \dots, \widehat{\tau_n^2}, \hat{\mu}, \widehat{\varphi^2}$  are numerically optimized by function **nlminb** in R. In detail, Hill Climbing was used to optimize these parameters for the iteration of each M-step, to efficiently cut down the time cost of function **nlminb**.

EM iterations terminated when reaching the convergence which was defined as no more than 0.01% changes on log-likelihood.