# Robust

Daniel Wells and Christopher Yau

August 4, 2016

## 1 Model

Let $\{y_i\}_{i=1}^n$ denote the coverage at position $i$ for $n$ loci across the chromosome/genome.
Let $\{s_i\}_{i=1}^n$ denote the segment number for each of those locations where $s_i \in \{1, \ldots, S\}$ and $S$ is the total number of segments.

$$p(y_i|s_i = j, w, \mu, \sigma^2) = (1-\rho_j) \underbrace{\sum_{c=1}^{C} w_{0,c} f(y_i; \mu_{0,c}, \sigma^2_{0,c})}_{\text{Common}} + \rho_j \underbrace{\sum_{k=1}^{K} w_{j,k} f(y_i; \mu_{j,k}, \sigma^2_{j,k})}_{\text{Segment-specific}} \tag{1}$$

where $f$ is the density function for the Normal distribution, $\sum_{k=1}^K w_{j,k} = 1$ and $\sum_{c=1}^C w_{0,c} = 1$.

## 2 EM Updates

### 2.1 E updates

For each data point $y_i$, $\psi_i$ gives the probability it is in a common (rather than segment specific) component.

$$\psi_i = \frac{(1-\rho_j) \sum_{c=1}^C w_{0,c} f(y_i; \mu_{0,c}, \sigma^2_{0,c})}{(1-\rho_j) \sum_{c=1}^C w_{0,c} f(y_i; \mu_{0,c}, \sigma^2_{0,c}) + \rho_j \sum_{k=1}^K w_{j,k} f(y_i; \mu_{j,k}, \sigma^2_{j,k})} \tag{2}$$

For each data point $y_i$, $\phi_{i,c}$ gives the probability it is in component $c$ given it's in a common component. $\nu_{i,k}$ gives the equivalent for segment specific components.

$$\phi_{i,c} = \frac{w_{0,c} f(y_i; \mu_{0,c}, \sigma^2_{0,c})}{\sum_{c=1}^C w_{0,c} f(y_i; \mu_{0,c}, \sigma^2_{0,c})},$$

$$\nu_{i,k} = \frac{w_{s_i,k} f(y_i; \mu_{s_i,k}, \sigma^2_{s_i,k})}{\sum_{k=1}^K w_{s_i,k} f(y_i; \mu_{s_i,k}, \sigma^2_{s_i,k})}$$

### 2.2 M updates

M update for $\rho$, a global common vs segment specific weighting:

$$\rho_j = 1 - \frac{\sum_{i:s_i=j}^n \psi_i}{n}$$

M updates for common component parameters:

$$w_{0,c} = \frac{\sum_{i=1}^n \psi_i \phi_{i,c}}{\sum_{i=1}^n \psi_i}$$

$$\mu_{0,c} = \frac{\sum_{i=1}^{n} \psi_i \phi_{i,c} y_i}{\sum_{i=1}^{n} \psi_i \phi_{i,c}}$$

$$\sigma_{0,c}^2 = \frac{\sum_{i=1}^{n} \psi_i \phi_{i,c} (y_i - \mu_{0,c})^2}{\sum_{i=1}^{n} \psi_i \phi_{i,c}}$$

M updates for segment specific component parameters:

$$w_{j,k} = \frac{\sum_{i:s_i=j} (1-\psi_i) \nu_{i,k}}{\sum_{i:s_i=j} (1-\psi_i)}$$

$$\mu_{j,k} = \frac{\sum_{i:s_i=j} (1-\psi_i) \nu_{i,k} y_i}{\sum_{i:s_i=j} (1-\psi_i) \nu_{i,k}}$$

$$\sigma_{j,k}^2 = \frac{\sum_{i:s_i=j} (1-\psi_i) \nu_{i,k} (y_i - \mu_{j,k})^2}{\sum_{i:s_i=j} (1-\psi_i) \nu_{i,k}}$$