EM Algorithm applied to Shim and Stephens (2015) - E-step

Brendan Law

2 February, 2019

We make slight modifications to the E-step of the EM algorithm that we derived in the HMT document (adapted from Crouse et al. (1998)), when applied to Shim and Stephens (2015). This is so that we can derive a computable, closed-form solution by using Bayes Factors (BF's).

1 Preamble

All preamble is as per the EM for Shim and Stephens (2015). We simplify the indexing of each node from s, l to i. Reinforcing previous derivations:

- At a given node of the tree, i, is this across ALL n individuals?

$$BF_i(y, g) := \frac{P(\boldsymbol{y}_i \mid g, \gamma_i = 1)}{P(\boldsymbol{y}_i \mid g, \gamma_i = 0)}$$

- We want to find the two quantities, for all m, n, i:

$$P(\gamma_i = m \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \tag{1}$$

$$P(\gamma_i = m, \gamma_{p(i)} = n \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \tag{2}$$

- $\mathcal{T}_i$ represents all wavelet coefficients at the subtree rooted at node i.
  - $\mathcal{T}_1$ represents all wavelet coefficients. (1 is the root of the tree).
  - $p(i)$ are the parents of i
  - $c(i)$ are the children of i
  - $p(i) \backslash i$ represents the tree rooted at $p(i)$, less the tree rooted at i
- From previously, we have:

$$\begin{aligned}
\beta_i(m) &:= P(\mathcal{T}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \Big[\prod_{j \in c(i)} P(\mathcal{T}_j \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})}) \Big] P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \Big[\prod_{j \in c(i)} \beta_{j,p(j)}(m) \Big] P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
\beta_{i,p(i)}(n) &:= P(\mathcal{T}_i \mid \gamma_{p(i)} = n, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \sum_{m=0}^{1} \beta_i(m) \epsilon_{i,p(i)}^{mn} \\
\beta_{p(i)\backslash i}(m) &:= P(\mathcal{T}_{p(i)\backslash i} \mid \gamma_{p(i)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \frac{\beta_{p(i)}(m)}{\beta_{i,p(i)}(m)} \\
\alpha_i(m) &:= P(\gamma_i = m, \mathcal{T}_{1\backslash i} \mid \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \sum_{n=0}^{1} \beta_{p(i)\backslash i}(n) \epsilon_{i,p(i)}^{mn} \alpha_{p(i)}(n)
\end{aligned}$$

- **New notation:** Let ν_i represent all *nodes* (as opposed to wavelet coefficients) in the subtree rooted at node i

2 Formal E-step algorithm

We denote the new (*'augmented'*) quantities with a superscript asterisk (*). The final amended algorithm is follows:

0. Initialise at **finest (lowest) scale**, $S = J$:

$$\beta_i^*(m) = \frac{P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} = BF_i(y, g) = \frac{\beta_i(m)}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})}$$

for each $m \in \{0, 1\}$.

1. $\forall \gamma_i$ at scale S, $\forall m \in \{0, 1\}$, calculate each of the following three quantities:

$$\beta_{i,p(i)}^*(m) = \sum_{n=0}^{1} \epsilon_{i,p(i)}^{nm} \beta_i^*(n) = \frac{\beta_{i,p(i)}(m)}{\prod_{j \in \nu_i} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \tag{3}$$

$$\begin{aligned} \beta_{p(i)}^*(m) &= \frac{P(\boldsymbol{y}_{p(i)} \mid \gamma_{p(i)} = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_{p(i)} \mid \gamma_{p(i)} = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \times \prod_{j \in c(p(i))} \beta_{j,p(j)}^*(m) \\ &= BF_{p(i)}(y, g) \times \prod_{j \in c(p(i))} \beta_{j,p(j)}^*(m) = \frac{\beta_{p(i)}(m)}{\prod_{j \in \nu_{p(i)}} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \end{aligned} \tag{4}$$

$$\beta_{p(i)\setminus i}^*(m) = \frac{\beta_{p(i)}^*(m)}{\beta_{i,p(i)}^*(m)} = \frac{\beta_{p(i)\setminus i}(m)}{\prod_{j \in \nu_{p(i)\setminus i}} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \tag{5}$$

2. $S := S - 1$

3. If $S = 1$ (coarsest/highest level), then stop, else return to step 1.

Down-step

0. Initialise state γ_1 at the **coarsest (highest)** scale level $S = 1$: $\alpha_1^*(m) = P(\gamma_1 = m, \mathcal{T}_{1\setminus 1} \mid \boldsymbol{\pi}^{(\mathrm{t})}) = P(\gamma_1 = m \mid \boldsymbol{\pi}^{(\mathrm{t})}) = P(\gamma_1 = m)$, for each $m \in \{0, 1\}$

1. $S := S + 1$

2. Calculate, $\forall \gamma_i$ at scale S, $\forall m \in \{0, 1\}$,

$$\alpha_i^*(m) = \sum_{n=0}^{1} \epsilon_{i,p(i)}^{mn} \alpha_{p(i)}^*(n) \beta_{p(i)\setminus i}^*(n) = \frac{\sum_{n=0}^{1} \epsilon_{i,p(i)}^{mn} \alpha_{p(i)}(n) \beta_{p(i)\setminus i}(n)}{\prod_{j \in \nu_{1\setminus i}} P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \tag{6}$$

3. If $S = J$ (finest/lowest level), then stop, else return to step 1.

From this, we can derive our two desired quantities. We see that, even though we have augmented the intermediate quantities slightly to use Bayes Factors, the cancelling means the results are equivalent to those desired from Crouse et al. (1998). For (1):

$$\begin{aligned} P(\gamma_i = m \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{\alpha_i^*(m)\beta_i^*(m)}{\sum_{n=0}^{1} \alpha_i^*(n)\beta_i^*(n)} \\ &= \frac{\alpha_i(m)\beta_i(m)}{\prod_{j \in \nu_1} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \frac{\prod_{j \in \nu_1} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})}{\sum_{n=0}^{1} \alpha_i(n)\beta_i(n)} \\ &= \frac{\alpha_i(m)\beta_i(m)}{\sum_{n=0}^{1} \alpha_i(n)\beta_i(n)} \end{aligned}$$

For (2):

$$\begin{aligned} P(\gamma_i = m, \gamma_{p(i)} = n \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{\beta_i^*(m)\epsilon_{i,p(i)}^{mn}\alpha_{p(i)}^*(n)\beta_{p(i)\setminus i}^*(n)}{\sum_{n=0}^{1} \alpha_i^*(n)\beta_i^*(n)} \\ &= \frac{\beta_i(m)\epsilon_{i,p(i)}^{mn}\alpha_{p(i)}(n)\beta_{p(i)\setminus i}(n)}{\prod_{j \in \nu_1} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \frac{\prod_{j \in \nu_1} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})}{\sum_{n=0}^{1} \alpha_i(n)\beta_i(n)} \\ &= \frac{\beta_i(m)\epsilon_{i,p(i)}^{mn}\alpha_{p(i)}(n)\beta_{p(i)\setminus i}(n)}{\sum_{n=0}^{1} \alpha_i(n)\beta_i(n)} \end{aligned}$$

DISREGARD THIS - WE'RE USING JUST ONE TREE. I'm guessing something like the below should still hold. (I don't see why it wouldn't) In the case where there are multiple trees (this corresponds, in our case, to multiple individuals, $i \in \{1, \ldots, n\}$, each individual represented by one tree), at each scale, we would repeat each of the iterations of the up and down algorithm n times, to find the desired quantities, resulting in (**??**) and (**??**) differing for each value of i:

$$P(\gamma_{sl}^{(i)} = m \mid \boldsymbol{y}^i, \boldsymbol{\pi}^{(\mathrm{t})}) = \frac{\beta_{sl}^{(i)}(m)\alpha_{sl}^{(i)}(m)}{\sum_{m=0}^{1} \beta_{sl}^{(i)}(m)\alpha_{sl}^{(i)}(m)}$$

$$P(\gamma_{sl}^{(i)} = m, \gamma_{p(sl)}^{(i)} = n \mid \boldsymbol{y}^i, \boldsymbol{\pi}^{(\mathrm{t})}) = \frac{\alpha_{p(sl)}^{(i)}(n)\beta_{p(sl)}^{(i)}(n)\beta_{sl}^{(i)}(m)\epsilon_{sl,p(sl)}^{mn\ (i)}}{\sum_{m=0}^{1} \beta_{sl}^{(i)}(m)\alpha_{sl}^{(i)}(m)}$$

3 Derivation of amended algorithm: (or run through brief example)

To show correctness, we focus on showing the correctness of (3) and (4), after step 0 is performed in the down-step. Showing the correctness of (5) and (6) is relatively straightforward after these two quantities are found, by using the augmented quantities instead of the original quantities.

For all nodes in the lowest scale, where $S = J$, we have:

$$\begin{aligned}
\beta_i^*(m) &= BF_i(y, g) \\
\therefore \beta_{i,p(i)}^*(m) &= \sum_{n=0}^{1} \epsilon_{i,p(i)}^{nm} \beta_i^*(n) \\
&= \sum_{n=0}^{1} P(\gamma_i = n \mid \gamma_{p(i)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) \frac{P(\boldsymbol{y}_i \mid \gamma_i = n, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\beta_{i,p(i)}(m)}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})}
\end{aligned}$$

Moving up the tree to the next highest scale, $S = J - 1$, we have:

$$\begin{aligned}
\beta_i^*(m) &= BF_i(y, g) \times \prod_{j \in c(i)} \beta_{j,p(j)}^*(m) \\
&= \frac{P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \times \prod_{j \in c(i)} \frac{\beta_{j,p(j)}(m)}{P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \times \prod_{j \in c(i)} \frac{P(\mathcal{T}_j \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{P(\mathcal{T}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{\prod_{j \in \nu_i} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})}
\end{aligned}$$

noting that the denominator is the product of each of i's children's individual wavelet coefficient densities, conditioned on each individual's $\gamma = 0$ (as opposed to a vector of coefficients as j is at the lowest scale) as well as from i itself. (At higher levels, the denominator becomes a product of the joint distribution of the coefficients in the subtrees rooted at each child. It remains a product, and not one big joint distribution of all these coefficients as each is conditioned on their own states, hence the joint distribution is not necessarily the product of the marginals as independence between these subtrees is not guaranteed). Note also that $p(j) = i$ as $j \in c(i)$. Moving back onto the $\beta_{i,p(i)}^*(m)$ term at this

scale:

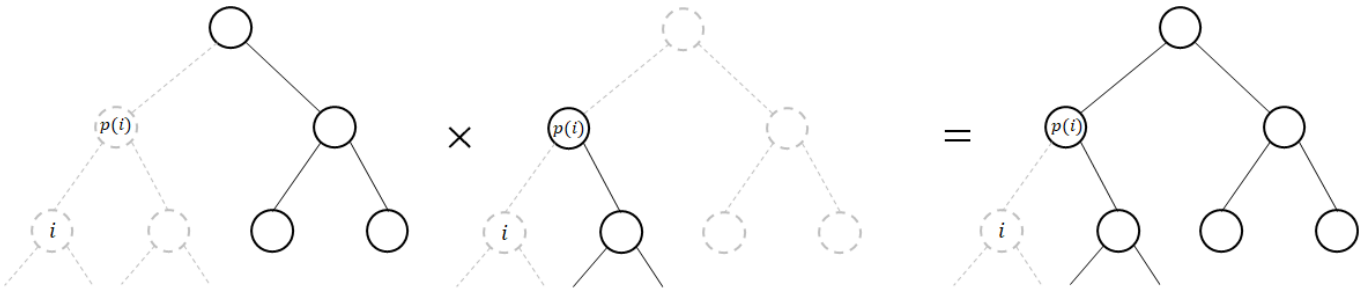
$$\begin{aligned}
\beta^*_{i,p(i)}(m) &= \sum_{n=0}^{1} \epsilon^{nm}_{i,p(i)} \beta^*_i(n) \\
&= \sum_{n=0}^{1} P(\gamma_i = n \mid \gamma_{p(i)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) \frac{P(\mathcal{T}_i \mid \gamma_i = n, \boldsymbol{\pi}^{(\mathrm{t})})}{\prod_{j \in \nu_i} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\beta_{i,p(i)}(m)}{\prod_{j \in \nu_i} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})}
\end{aligned}$$

Lastly, we'll move up the tree one more scale level, $S = J - 2$, and confirm that the denominator of $\beta^*_i(m)$ does remain as a product of the density of individual wavelet coefficients contained in the subtree rooted at i, in the general case.

$$\begin{aligned}
\beta^*_i(m) &= BF_i(y, g) \times \prod_{j \in c(i)} \beta^*_{j,p(j)}(m) \\
&= \frac{P(\boldsymbol{y}_i \mid \gamma_i = m, \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \times \prod_{j \in c(i)} \frac{\beta_{j,p(j)}(m)}{\prod_{k \in \nu_j} P(\boldsymbol{y}_k \mid \gamma_k = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\beta_i(m)}{P(\boldsymbol{y}_i \mid \gamma_i = 0, \boldsymbol{\pi}^{(\mathrm{t})}) \prod_{j \in c(i)} \prod_{k \in \nu_j} P(\boldsymbol{y}_k \mid \gamma_k = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \qquad \text{(numerator result from the non-augmented derivations)} \\
&= \frac{\beta_i(m)}{\prod_{j \in \nu_i} P(\boldsymbol{y}_j \mid \gamma_j = 0, \boldsymbol{\pi}^{(\mathrm{t})})} \qquad \text{(denominator as required)}
\end{aligned}$$

where the denominator in the last step follows as, for each of i's children, $j \in c(i)$, all the nodes in each subtree rooted at each j, ν_j, are mutually exclusive. Hence the product of all the terms in the denominator results in the product of densities of the individual wavelet coefficients contained in the subtree rooted at i, all the way up to the highest level.

(5) follows due to cancellation between the overlapping products of the individual densities at the subtrees rooted at $p(i)$ and i. (6) follows as the separate products in the denominators from the augmented α and β terms are mutually exclusive, and constitute the whole tree, minus the subtree rooted at i. In other words, the denominator of the α term has WCs from $\nu_{1 \setminus p(i)}$, and, β's denominators are the WCs from the $\nu_{p(i) \setminus i}$ terms. Hence, multiplied together, we get the probabilities from the $\nu_{1 \setminus i}$ WCs.



References

M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on signal processing*, 46(4):886–902, 1998.

B. Servin and M. Stephens. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics*, 3(7):e114, 2007.

H. Shim and M. Stephens. Wavelet-based genetic association analysis of functional phenotypes arising from high-throughput sequencing assays. *The annals of applied statistics*, 9(2):655, 2015.