EM Algorithm applied to Shim and Stephens (2015)

Brendan Law

11 January, 2019

A document showing my formal(-ish) derivation for the Expectation Maximisation (EM) algorithm as applied to Shim and Stephens (2015). The task here is to parameterise the probability that our wavelet coefficients come from either one of two latent states, each with their own conditional densities. A lot of quantities are divided by a constant to convert them into Bayes Factors, for convenience, as the closed form of the underlying conditional densities are less convenient to work with.

1 Preamble

We are working with data which has been transformed into wavelet coefficients (WC)'s across various states and scales, as well as dependent variables (eg: SNPs of interest) which we are trying to regress onto the WCs to perform association analysis across individuals, and eventually, across groups.

- Individuals, $i \in \{1, \ldots, n\}$
- Scales, $s \in \{1, \ldots, J\}$
- Locations, $l \in \{1, \ldots, L_s\}$ (as each scale has a different number of locations, ranging from one location at the coarsest scale to many locations at the finest scale)
- Note that although Crouse et al. (1998) refers to $s = J$ as the coarsest scale, and $s = 1$ as the finest scale, our notation will adopt *the opposite;* $s = J$ is the finest scale (bottom) and $s = 1$ is the coarset scale (top), and hence location $1, 1$ refers to the wavelet coefficient at the only location at the coarsest scale.
- For each scale and location, s, l, a vector of WCs: $\boldsymbol{y}_{sl} = (y^1_{sl}, \ldots, y^n_{sl})$, and $\boldsymbol{Y} = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{JL_J})$
- A vector of binary indicator variables used to indicate whether $\boldsymbol{y}_{sl}$ is associated with g (1 for association): $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{JL_J})$, where $\gamma_{sl} \in \{0, 1\}$
  - This is the latent state variable in this setting
- Parameter set, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$
  - These are the free parameters we will be parameterising through the EM algorithm
- Dependent variables, $\boldsymbol{g} = (g^1, \ldots, g^n)$, which represents a vector of data points, one from each individual - in this paper, the genotype data (number of copies of the minor allele) for individual i at a single SNP of interest.
- Hierarchical model:
  - $y^i_{sl} = \mu_{sl} + \gamma_{sl}\beta_{sl}g^i + \epsilon^i_{sl}$ with $\epsilon^i_{sl} \sim \mathcal{N}(0, \sigma^2_{sl})$, implying:
    * $P(y^i_{sl} \mid \mu_{sl}, \gamma_{sl} = 0, \beta_{sl}, g^i, \sigma^2_{sl}) \sim \mathcal{N}(\mu_{sl}, \sigma^2_{sl})$, and
    * $P(y^i_{sl} \mid \mu_{sl}, \gamma_{sl} = 1, \beta_{sl}, g^i, \sigma^2_{sl}) \sim \mathcal{N}(\mu_{sl} + \beta_{sl}g^i, \sigma^2_{sl})$
  - $P(\gamma_{sl} = 1 \mid \boldsymbol{\pi}) = \pi_s$ for each scale s, across all locations, l

Some extra notation and assumptions regarding the model:

- Note that given their state, γ_{sl}, WCs are conditionally independent across scales and locations
- Due to the Bayesian setting of this model, the π_s's are hyperparameters, not random variables
- $P(\gamma_{sl} = 1 \mid \boldsymbol{\pi}) = \pi_s$ and $P(\gamma_{sl} = 0 \mid \boldsymbol{\pi}) = 1 - \pi_s$

- $\pi_s = 0 \Rightarrow \gamma_{sl} = 0$ and consequently $\boldsymbol{\pi} \equiv 0 \Rightarrow \boldsymbol{\gamma} \equiv 0$
- The Bayes Factor is used extensively in the paper to measure the support for $\gamma_{sl} = 1$, for a specific s, l, across all individuals i. It is easier to compute (has a closed form) thanks to the models and priors from Servin and Stephens (2007). It is denoted as such:

$$\mathrm{BF}_{sl}(y, g) := \frac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 1, \boldsymbol{g})}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})}$$

2 Complete log likelihood derivation

$$\begin{aligned}
P(\boldsymbol{Y}, \boldsymbol{\gamma} \mid \boldsymbol{g}, \boldsymbol{\pi}) &= P(\boldsymbol{Y} \mid \boldsymbol{\gamma}, \boldsymbol{g}, \boldsymbol{\pi}) P(\boldsymbol{\gamma} \mid \boldsymbol{g}, \boldsymbol{\pi}) \\
&= P(\boldsymbol{Y} \mid \boldsymbol{\gamma}, \boldsymbol{g}, \boldsymbol{\pi}) \prod_{s,l} P(\gamma_{sl} \mid \boldsymbol{\pi}) && \text{(independence of } \pi_s \text{ across scales, } \gamma \text{ independent of } \boldsymbol{g}) \\
&= P((\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{JL}) \mid \boldsymbol{\gamma}, \boldsymbol{g}) \prod_{s,l} P(\gamma_{sl} \mid \boldsymbol{\pi}) && \text{(independence of } \boldsymbol{y} \text{ of } \boldsymbol{\pi}) \\
&= \prod_{s,l} \left[P(\boldsymbol{y}_{sl} \mid \gamma_{sl}, \boldsymbol{g}) P(\gamma_{sl} \mid \pi_s) \right] && \text{(independence of } \boldsymbol{y}_{sl}\text{'s conditional on own state)} \\
&= \prod_{s,l} \prod_{k=0}^{1} \left[P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k, \boldsymbol{g}) P(\gamma_{sl} = k \mid \pi_s) \right]^{\mathbb{1}\{\gamma_{sl} = k\}} \\
&= P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) \prod_{s,l} \prod_{k=0}^{1} \frac{\left[P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k, \boldsymbol{g}) P(\gamma_{sl} = k \mid \pi_s) \right]^{\mathbb{1}\{\gamma_{sl} = k\}}}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})}
\end{aligned}$$

with the last step due to:

$$P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) = \prod_{s,l} \prod_{k=0}^{1} P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})$$

Therefore, the complete log likelihood:

$$\begin{aligned}
\log L(\boldsymbol{\pi}; \boldsymbol{Y}, \boldsymbol{\gamma}, \boldsymbol{g}) &= \log P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) + \sum_{s,l} \Bigg[\mathbb{1}\{\gamma_{sl} = 0\} \Bigg(\log \frac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})} + \log(1 - \pi_s) \Bigg) \ldots \\
&\quad + \mathbb{1}\{\gamma_{sl} = 1\} \Bigg(\log \frac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 1, \boldsymbol{g})}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0, \boldsymbol{g})} + \log \pi_s \Bigg) \Bigg] \\
&= \log P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) + \sum_{s,l} \Big[\mathbb{1}\{\gamma_{sl} = 0\} \big(\log(1 - \pi_s) \big) + \mathbb{1}\{\gamma_{sl} = 1\} \big(\log \mathrm{BF}_{sl}(y, g) + \log \pi_s \big) \Big]
\end{aligned}$$

Note that, as always, the following remains a random variable representing the unknown state of the γ_{sl} variable:

$$\mathbb{1}\{\gamma_{sl} = k\} = \begin{cases} 1 & \gamma_{sl} = k \\ 0 & \gamma_{sl} \neq k \end{cases}$$

3 EM algorithm

We will now compute the MLE of the parameters in $\boldsymbol{\pi}$ by iterating through the EM algorithm and updating $\boldsymbol{\pi}$ at the end of each step.

$$\begin{aligned}
Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})}) &= \mathbb{E}_{(\boldsymbol{\gamma} \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})})} [\log L(\boldsymbol{\pi}; \mathbf{Y}, \boldsymbol{\gamma}, \mathbf{g})] && \text{(finding unknowns in this statement is the E-step)} \\
\boldsymbol{\pi}^{(\mathrm{t+1})} &:= \operatorname*{argmax}_{\boldsymbol{\pi}} Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})}) && \text{(this is the M-step)}
\end{aligned}$$

4 E-step: derivation

$$
\begin{aligned}
Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})}) &= \mathbb{E}_{(\boldsymbol{\gamma}|\boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}[\log L(\boldsymbol{\pi};\boldsymbol{Y},\boldsymbol{\gamma},\boldsymbol{g})] \\
&= \mathbb{E}_{(\boldsymbol{\gamma}|\boldsymbol{Y},\boldsymbol{g}\boldsymbol{\pi}^{(\mathrm{t})})}\Big[\log P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) + \sum_{s,l}\big[\mathbb{1}\{\gamma_{sl} = 0\}\big(\log(1-\pi_s)\big)\ldots \\
&+ \mathbb{1}\{\gamma_{sl} = 1\}\big(\log \mathrm{BF}_{sl}(y,g) + \log \pi_s\big)\big]\Big] \\
&= \log P(\boldsymbol{Y} \mid \boldsymbol{\gamma} \equiv 0, \boldsymbol{g}) + \sum_{i=1}^{n}\sum_{s,l}\big[P(\gamma_{sl} = 0 \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})\big(\log(1-\pi_s)\big) + P(\gamma_{sl} = 1 \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})\big(\log \mathrm{BF}_{sl}(y,g) + \log \pi_s\big)\big]
\end{aligned}
$$

Now we can evaluate each of the two conditional probability statements around γ_{sl}:

$$
\begin{aligned}
P(\gamma_{sl} = k \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})}) &= P(\gamma_{sl} = k \mid \boldsymbol{y}_{sl},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})}) && \text{(independent of WCs from other scales, locations)} \\
&= \frac{P(\gamma_{sl} = k, \boldsymbol{y}_{sl} \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}{\sum_{k'} P(\gamma_{sl} = k', \boldsymbol{y}_{sl} \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k,\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}{\sum_{k'} P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k',\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k' \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k,\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}{\sum_{k'} P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k',\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k' \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\dfrac{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k,\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0,\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}}{\dfrac{\sum_{k'} P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = k',\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})P(\gamma_{sl} = k' \mid \boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{y}_{sl} \mid \gamma_{sl} = 0,\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})}} \\
& && \text{(Divide both sides by a constant to convert into Bayes Factors)} \\
\therefore P(\gamma_{sl} = 1 \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{\mathrm{BF}_{sl}(y,g)\pi_s^{(\mathrm{t})}}{BF_{sl}(y,g)\pi_s^{(\mathrm{t})} + (1-\pi_s^{(\mathrm{t})})}, \text{ and} \\
\therefore P(\gamma_{sl} = 0 \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{1-\pi_s^{(\mathrm{t})}}{BF_{sl}(y,g)\pi_s^{(\mathrm{t})} + (1-\pi_s^{(\mathrm{t})})}
\end{aligned}
$$

5 M-step: maximisation

To simplify the notation, denote:

$$A_{sl,k}^{(\mathrm{t})} := P(\gamma_{sl} = k \mid \boldsymbol{Y},\boldsymbol{g},\boldsymbol{\pi}^{(\mathrm{t})})$$

We have that:

$$\boldsymbol{\pi}^{(\mathrm{t+1})} := \underset{\boldsymbol{\pi}}{\operatorname{argmax}}\, Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})})$$

where:

$$Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})}) \propto \sum_{s,l}\big[A_{sl,0}^{(\mathrm{t})}\log(1-\pi_s) + A_{sl,1}^{(\mathrm{t})}(\log BF_{sl}(y,g) + \log \pi_s)\big]$$

(proportionality up to the constant where $\boldsymbol{\gamma} \equiv 0$ up front)

Hence, finding the $\boldsymbol{\pi}$ which maximises the term in the sum on the right hand side above will yield an equivalent result as the one desired.

For each $s \in \{1, \dots, J\}$,

$$\begin{aligned}\frac{\delta Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})})}{\delta \pi_s} &= \sum_{l=1}^{L_s} \Big(- \frac{A_{sl,0}^{(\mathrm{t})}}{1-\pi_s} + \frac{A_{sl,1}^{(\mathrm{t})}}{\pi_s} \Big) \\ &= \frac{\sum_{l=1}^{L_s} (-A_{sl,0}^{(\mathrm{t})} \pi_s + A_{sl,1}^{(\mathrm{t})} (1-\pi_s))}{\pi_s(1-\pi_s)}\end{aligned}$$

Setting this equal to zero,

$$\begin{aligned}0 &= \frac{\sum_{l=1}^{L_s} (-A_{sl,0}^{(\mathrm{t})} \pi_s^{(\mathrm{t+1})} + A_{sl,1}^{(\mathrm{t})} (1-\pi_s^{(\mathrm{t+1})}))}{\pi_s^{(\mathrm{t+1})}(1-\pi_s^{(\mathrm{t+1})})} \\ \Rightarrow \sum_{l=1}^{L_s} A_{sl,0}^{(\mathrm{t})} \pi_s^{(\mathrm{t+1})} &= \sum_{l=1}^{L_s} A_{sl,1}^{(\mathrm{t})} (1-\pi_s^{(\mathrm{t+1})}) \\ \pi_s^{(\mathrm{t+1})} \sum_{l=1}^{L_s} A_{sl,0}^{(\mathrm{t})} &= (1-\pi_s^{(\mathrm{t+1})}) \sum_{l=1}^{L_s} A_{sl,1}^{(\mathrm{t})} \\ (\pi_s^{(\mathrm{t+1})}) \sum_{l=1}^{L_s} (A_{sl,0}^{(\mathrm{t})} + A_{sl,1}^{(\mathrm{t})}) &= \sum_{l=1}^{L_s} A_{sl,1}^{(\mathrm{t})} \\ \Rightarrow \pi_s^{(\mathrm{t+1})} &= \frac{\sum_{l=1}^{L_s} A_{sl,1}^{(\mathrm{t})}}{\sum_{l=1}^{L_s} (A_{sl,0}^{(\mathrm{t})} + A_{sl,1}^{(\mathrm{t})})} = \frac{\sum_{l=1}^{L_s} A_{sl,1}^{(\mathrm{t})}}{L_s}\end{aligned}$$

for each scale, s. The simplification in the last line follows as:

$$\sum_{l=1}^{L_s} (A_{sl,0}^{(\mathrm{t})} + A_{sl,1}^{(\mathrm{t})}) = \sum_{l=1}^{L_s} 1 = L_s$$

6 Extension to HMT structure

Now we extend the above model by imposing a Hidden Markov Tree (HMT) structure on the states, rather than assuming they are independent across scales and locations (see the 'EM_algo_HMT.pdf' file for more details behind the generic derivation).

6.1 Changed assumptions:

- $P(\gamma_{sl} = 1 \mid \boldsymbol{\pi}) = \pi_s$ across all locations, l, for each scale, s. Instead, γ_{sl} is now governed by a HMT tree structure.
- The parameter now required for this extension now (still denoted by $\boldsymbol{\pi}$) contains:
  - $\pi_{sl}^k = P(\gamma_{sl} = k \mid \boldsymbol{\pi})$
  - $\epsilon_{sl,p(sl)}^{kl} = P(\gamma_{sl} = k \mid \gamma_{p(sl)} = l, \boldsymbol{\pi})$
- With constraints:
  - $\sum_{k=0}^{1} \pi_{sl}^k = 1$
  - $\sum_{k=0}^{1} \epsilon_{sl,p(sl)}^{kl} = 1$
- Note that, for each tree, we really only need:
  - π_{11}^k (the param of the root node), and
  - $\epsilon_{sl,p(sl)}^{kl}$ for $s \in \{2, \dots, J\}$ and $l \in \{1, \dots, L_s\}$ for $k \in \{0, 1\}$,
  - to fully parameterise all the probabilities - the tree 'root' probabilities plus the 'transition' probabilities will be sufficient to generate probabilities of all states and locations.

6.2 Complete likelihood derivation

$$
\begin{aligned}
P(\mathbf{Y}, \boldsymbol{\gamma} \mid \boldsymbol{g}, \boldsymbol{\pi}) &= P(\mathbf{Y} \mid \boldsymbol{\gamma}, \boldsymbol{g}, \boldsymbol{\pi}) P(\boldsymbol{\gamma} \mid \boldsymbol{g}, \boldsymbol{\pi}) \\
&= \left[\prod_{s,l} P(\boldsymbol{y}_{sl} \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi}) \right] P(\gamma_{11}, \ldots, \gamma_{JL_s} \mid \boldsymbol{g}, \boldsymbol{\pi}) \\
&= \left[\prod_{s,l} P(\boldsymbol{y}_{sl} \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi}) \right] P(\gamma_{11} \mid \boldsymbol{g}, \boldsymbol{\pi}) \prod_{s=2}^{J} \prod_{l=1}^{L_s} P(\gamma_{sl} \mid \gamma_{p(sl)}, \boldsymbol{g}, \boldsymbol{\pi}) P(\gamma_{11}, \ldots, \gamma_{JL_t s} \mid \boldsymbol{g}, \boldsymbol{\pi}) \\
&\qquad \text{(where the second term comes from the HMT derivations)} \\
&= \prod_{s,l} \prod_{m=0}^{1} P(\boldsymbol{y}_{sl} = m \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi})^{\mathbb{1}\{\gamma_{sl} = m\}} \prod_{m=0}^{1} P(\gamma_{11} = m \mid \boldsymbol{g}, \boldsymbol{\pi})^{\mathbb{1}\{\gamma_{11} = m\}} \ldots \\
&\times \prod_{s=2}^{J} \prod_{l=1}^{L_s} \prod_{m=0}^{1} \prod_{n=0}^{1} P(\gamma_{sl} = m \mid \gamma_{p(sl)} = n, \boldsymbol{g}, \boldsymbol{\pi})^{\mathbb{1}\{\gamma_{sl} = m\} \mathbb{1}\{\gamma_{p(sl)} = n\}} \\
&= \prod_{s,l} \prod_{m=0}^{1} P(\boldsymbol{y}_{sl} = m \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi})^{\mathbb{1}\{\gamma_{sl} = m\}} \prod_{m=0}^{1} (\pi_{11}^{m})^{\mathbb{1}\{\gamma_{11} = m\}} \prod_{s=2}^{J} \prod_{l=1}^{L_s} \prod_{m=0}^{1} \prod_{n=0}^{1} (\epsilon_{sl,p(sl)}^{mn})^{\mathbb{1}\{\gamma_{sl} = m\} \mathbb{1}\{\gamma_{p(sl)} = n\}} \\
&\Rightarrow \log L(\boldsymbol{\pi}; \mathbf{Y}, \boldsymbol{\gamma}, \boldsymbol{g}) = \sum_{s,l} \sum_{m=0}^{1} \mathbb{1}\{\gamma_{sl} = m\} \log P(\boldsymbol{y}_{sl} = m \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi}) \ldots \\
&\qquad + \mathbb{1}\{\gamma_{11} = 0\} \log(\pi_{11}^{0}) + \mathbb{1}\{\gamma_{11} = 1\} \log(\pi_{11}^{1}) \ldots \\
&\qquad + \sum_{s=2}^{J} \sum_{l=1}^{L_s} \sum_{m=0}^{1} \sum_{n=0}^{1} \mathbb{1}\{\gamma_{sl} = m\} \mathbb{1}\{\gamma_{p(sl)} = n\} \log(\epsilon_{sl,p(sl)}^{mn})
\end{aligned}
$$

6.3 E-step

$$
\begin{aligned}
Q(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(\mathrm{t})}) &= \mathbb{E}_{(\boldsymbol{\gamma} \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})})}[\log L(\boldsymbol{\pi}; \mathbf{Y}, \boldsymbol{\gamma}, \mathbf{g})] \\
&= \sum_{s,l} \sum_{m=0}^{1} P(\gamma_{sl} = m \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \log P(\boldsymbol{y}_{sl} = m \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi}) \ldots \\
&+ P(\gamma_{11} = 0 \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\pi_{11}^{0}) + P(\gamma_{11} = 1 \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\pi_{11}^{1}) \ldots \\
&+ \sum_{s=2}^{J} \sum_{l=1}^{L_s} \sum_{m=0}^{1} \sum_{n=0}^{1} P(\gamma_{sl} = m, \gamma_{p(sl)} = n \mid \boldsymbol{Y}, \boldsymbol{g}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\epsilon_{sl,p(sl)}^{mn}) \\
&= \sum_{s,l} \sum_{m=0}^{1} P(\gamma_{sl} = m \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) \log P(\boldsymbol{y}_{sl} = m \mid \gamma_{sl}, \boldsymbol{g}, \boldsymbol{\pi}) \ldots \\
&+ P(\gamma_{11} = 0 \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\pi_{11}^{0}) + P(\gamma_{11} = 1 \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\pi_{11}^{1}) \ldots \\
&+ \sum_{s=2}^{J} \sum_{l=1}^{L_s} \sum_{m=0}^{1} \sum_{n=0}^{1} P(\gamma_{sl} = m, \gamma_{p(sl)} = n \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) \log(\epsilon_{sl,p(sl)}^{mn})
\end{aligned}
$$

The resulting quantities to be solved for are:

$$
\begin{aligned}
P(\gamma_{sl} = m \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{P(\gamma_{sl} = m, \boldsymbol{Y} \mid \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{Y} \mid \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\beta_{sl}(m)\alpha_{sl}(m)}{\displaystyle\sum_{m=0}^{1} \beta_{sl}(m)\alpha_{sl}(m)} \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
P(\gamma_{sl} = m, \gamma_{p(sl)} = n \mid \boldsymbol{Y}, \boldsymbol{\pi}^{(\mathrm{t})}) &= \frac{P(\gamma_{sl} = m, \gamma_{p(sl)} = n, \boldsymbol{Y} \mid \boldsymbol{\pi}^{(\mathrm{t})})}{P(\boldsymbol{Y} \mid \boldsymbol{\pi}^{(\mathrm{t})})} \\
&= \frac{\alpha_{p(sl)}(n)\beta_{p(sl)}(n)\beta_{sl}(m)\epsilon_{sl,p(sl)}^{mn}}{\displaystyle\sum_{m=0}^{1} \beta_{sl}(m)\alpha_{sl}(m)} \qquad (2)
\end{aligned}
$$

for $m \in \{0,1\}, n \in \{0,1\}, s \in \{1,\ldots,J\}, l \in \{1,\ldots,L_s\}$. The remaining terms are parameterised as part of $\boldsymbol{\pi}^{(\mathrm{t})}$. The resulting quantities (1) and (2) rely on evaluating the below (derivations from the HMT analysis):

$$
\begin{aligned}
\beta_{sl}(m) &:= P(\mathcal{T}_{sl} \mid \gamma_{sl} = m, \boldsymbol{\pi}^{(\mathrm{t})}) && (3) \\
\beta_{p(sl)}(m) &:= P(\mathcal{T}_{p(sl)} \mid \gamma_{p(sl)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \Big[\prod_{j \in c(p(sl))} P(\mathcal{T}_j \mid \gamma_{p(sl)} = m, \boldsymbol{\pi}^{(\mathrm{t})})\Big] P(\boldsymbol{y}_{p(sl)} \mid \gamma_{p(sl)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) && (4) \\
\beta_{sl,p(sl)}(n) &:= P(\mathcal{T}_{sl} \mid \gamma_{p(sl)} = n, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \sum_{m=0}^{1} \beta_{sl}(m)\epsilon_{sl,p(sl)}^{mn} && (5) \\
\beta_{p(sl)\setminus sl}(m) &:= P(\mathcal{T}_{p(sl)\setminus sl} \mid \gamma_{p(sl)} = m, \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \frac{\beta_{p(sl)}(m)}{\beta_{sl,p(sl)}(m)} && (6) \\
\alpha_{sl}(m) &:= P(\gamma_{sl} = m, \mathcal{T}_{11\setminus sl} \mid \boldsymbol{\pi}^{(\mathrm{t})}) \\
&= \sum_{n=0}^{1} \beta_{p(sl)\setminus sl}(n)\epsilon_{sl,p(sl)}^{mn}\alpha_{p(sl)}(n) && (7)
\end{aligned}
$$

where $\mathcal{T}_{sl}$ is the vector of wavelet coefficients in the tree rooted at scale s and location l, and $\mathcal{T}_{11}$ is the complete vector of wavelet coefficients for a given individual. These values are found by applying the upward-downward algorithm as follows:

Up-step

0. Initialise at **finest (lowest) scale**, $S = J$: $\beta_{sl}(m) = f(\boldsymbol{y}_{sl} \mid \gamma_{sl} = m, \boldsymbol{\pi}^{(\mathrm{t})})$ for each $m \in \{0,1\}$
1. $\forall \gamma_{sl}$ at scale S, $\forall m \in \{0,1\}$, calculate each of the following three quantities:
   - $\beta_{sl,p(sl)}(m)$
   - $\beta_{p(sl)}(m)$
   - $\beta_{p(sl)\setminus sl}(m)$
2. $S := S - 1$
3. If $S = 1$ (coarsest/highest level), then stop, else return to step 1.

Down-step

0. Initialise state γ_{11} at the **coarsest (highest)** scale level $S = 1$: $\alpha_{sl}(m) = P(\gamma_{11} = m, \mathcal{T}_{11\setminus 11} \mid \boldsymbol{\pi}^{(\mathrm{t})}) = P(\gamma_{11} = m \mid \boldsymbol{\pi}^{(\mathrm{t})}) = P(\gamma_{11} = m)$, for each $k \in \{0,1\}$
1. $S := S + 1$

2. Calculate, $\forall \gamma_{sl}$ at scale S, $\forall m \in \{0,1\}$, $\alpha_{sl}(m)$

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

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 (1) and (2) differing for each value of i:

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

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

I am a little confused right here as to how to incorporate Bayes Factors into this. A lot of the quantities above we can't solve in closed form, and although dividing by the $P(\boldsymbol{Y} \mid \gamma \equiv 0)$ constant usually works, it doesn't quite work out for quantity (4), which is a product of terms. I think for everything else, it probably cancels out on both the numerator and denominator somehow, like it did in Ash.

6.4 M-step

Once we've found the required quantities, given multiple (N) trees, we can compute the maximising quantities of $\boldsymbol{\pi}$ as follows.

Denote

$$A_{sl}^i(m) := P(\gamma_{sl}^{(i)} = m \mid \boldsymbol{y}^i, \boldsymbol{\pi}^{(\mathrm{t})})$$
$$B_{sl,p(sl)}^i(mn) := P(\gamma_{sl}^{(i)} = m, \gamma_{p(sl)}^{(i)} = n \mid \boldsymbol{y}^i, \boldsymbol{\pi}^{(\mathrm{t})})$$

Then, for each s, l and states m, n:

$$(\pi_{sl}^m)^{(t+1)} = \frac{\sum_{i=1}^N A_{sl}^i(m)}{N}, \text{ and}$$
$$(\epsilon_{sl,p(sl)}^{mn})^{(t+1)} = \frac{\sum_{i=1}^N B_{sl,p(sl)}^i(mn)}{\sum_{i=1}^N A_{p(sl)}^i(n)}$$

The derivations for these are found in the HMT documentation.

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.