Dynamic algorithm for WaveQTL HMT – posterior means and variances

Brendan Law

October 6, 2019

1 DP algorithms for approximating (pointwise) posterior quantities

Consider a tree with 3 scales, 7 nodes (as per the example in the thesis). The idea is to calculate things in adjacent pairs, and to exploit their conditional independence, along with their relationships to their parents. Also, we know, for $\in \{0, 1\}$ (note i've dropped the $\widehat{\theta}$ for brevity):

$$P(\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl} = m) \tag{1}$$
$$E(\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl} = m) \tag{2}$$
$$V(\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl} = m) \tag{3}$$

and from the EM algorithm:

$$P(\gamma_{sl} = m \mid \boldsymbol{y}, \boldsymbol{g}) \tag{4}$$
$$P(\gamma_{sl} = n \mid \gamma_{p(sl)} = m, \boldsymbol{y}, \boldsymbol{g}) \tag{5}$$

For the last two quantities, similarly to WaveQTL without HMT, we have:

$$\begin{aligned}
E[\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}] &= E[E[\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl}]] && (6)\\
&= \sum_{m=0}^{1} P(\gamma_{sl} = m \mid \boldsymbol{y}, \boldsymbol{g}) E[\beta_{sl} \mid y_{sl}, \boldsymbol{g}, \gamma_{sl} = m] && (7)\\
&= \phi_{sl} a_{sl} && (8)\\
V(\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}) &= E[\beta_{sl}^2 \mid \boldsymbol{y}, \boldsymbol{g}] - E[\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}]^2 && (9)\\
&= E[E[\beta_{sl}^2 \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl}]] - (\sum_{m=0}^{1} \phi_{sl} a_{sl})^2 && (10)\\
&= \sum_{m=0}^{1} P(\gamma_{sl} = m \mid \boldsymbol{y}, \boldsymbol{g})(V(\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl} = m) + E[\beta_{sl} \mid \boldsymbol{y}, \boldsymbol{g}, \gamma_{sl} = m]^2) - (\phi_{sl} a_{sl})^2 && (11)\\
&= \phi_{sl}(\nu_{sl} + a_{sl}^2) - (\phi_{sl} a_{sl})^2 && (12)
\end{aligned}$$

Now, the tricky bit starts when we want to find the data space quantities as we need to sum across all scales and locations, which are no longer guaranteed to be independent of each other. This causes an issue for variances, but not for means. Means can actually be found using the WaveQTL method, but we demonstrate how to do the DP version here as its quantities are required to compute the variances. Here, we demonstrate the algorithm required to calculate the data-space means and variances at one base. However, in practice, the calculations across each base can be performed independently of each other, allowing a vector of quantities to be updated through the algorithm at once, rather than repeating this algorithm B times.

Here is some further notation. Denote $(j, c_1), (j, c_2)$ as the children of the node at $(j-1, l)$. Then:

$$\mu_{j-1,l}(\gamma_{j,c_1}, \gamma_{j,c_2}, k) := E[w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k, \gamma_{j,c_1}, \gamma_{j,c_2}] \tag{13}$$
$$\sigma^2_{j-1,l}(\gamma_{j,c_1}, \gamma_{j,c_2}, k) := V(w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k, \gamma_{j,c_1}, \gamma_{j,c_2}) \tag{14}$$

Also, denote all nodes in the subtree rooted at (s,l) as $\nu(s,l)$, and:

$$E[\mathcal{T}_{s,l}] = E[\sum_{j\in\nu(s,l)} w_j\beta_j] \tag{15}$$

$$V(\mathcal{T}_{s,l}) = V(\sum_{j\in\nu(s,l)} w_j\beta_j) \tag{16}$$

1.1 Posterior mean

The quantity we want to calculate is:

$$E[\alpha_b \mid \boldsymbol{y},\boldsymbol{g}] = E[\sum_{s,l} w_{(s,l),b}\beta_{sl} \mid \boldsymbol{y},\boldsymbol{g}] \tag{17}$$

which obviously, we can evaluate, using additivity of means, despite the dependence between quantities at different scales and locations. To simplify notation, we drop the conditioning on $\boldsymbol{y},\boldsymbol{g}$. The algorithm is as follows:

0. **Initialise:** At the finest scale, $j = J$ (bottom of tree), for state $m = 0, 1$, evaluate, at all locations:

$$\mathbb{E}[\beta_{J,l} \mid \gamma_{J,l} = m] \tag{18}$$

   Note that in practice, this will be 0 when $m = 0$.

1. At scale j, note that elements at scale $j-1$ are parents of those in scale j. Calculate, for all locations, for all states, $k = 0, 1$:

$$\begin{aligned}
E[\mathcal{T}_{j-1,l} \mid \gamma_{j-1,l} = k] &= E[w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k] &(19)\\
&= E[E[w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j,c_1},\gamma_{j,c_2},\gamma_{j-1,l} = k]] &(20)\\
&= E[\mu_{j-1,l}(\gamma_{j,c_1},\gamma_{j,c_2},k)] &(21)\\
&= \sum_{a,m} P(\gamma_{j,c_1} = a,\gamma_{j,c_2} = m \mid \gamma_{j-1,l} = k)\mu_{j-1,l}(\gamma_{j,c_1} = a,\gamma_{j,c_2} = m,k) &(22)\\
&= \sum_{a=0}^{1}\sum_{m=0}^{1} P(\gamma_{j,c_1} = a \mid \gamma_{j-1,l} = k)P(\gamma_{j,c_2} = m \mid \gamma_{j-1,l} = k)\mu_{j-1,l}(\gamma_{j,c_1} = a,\gamma_{j,c_2} = m,k) &(23)
\end{aligned}$$

   due to independence of the state variables when conditioned on the parent state. By additivity of means:

$$\mu_{j-1,l}(\gamma_{j,c_1} = a,\gamma_{j,c_2} = m,k) = w_{j-1,l}E[\beta_{j-1,l} \mid \gamma_{j-1,l} = k] + E[\mathcal{T}_{j,c_1} \mid \gamma_{j,c_1} = a] + E[\mathcal{T}_{j,c_2} \mid \gamma_{j,c_2} = m] \tag{24}$$

   where the second and third terms have been evaluated in previous iterations of this algorithm.

2. Set $j := j - 1$

3. If $j = 1$ (coarsest/least granular level), then stop, else return to step 1.

4. To get the desired quantity, marginalise over all states:

$$\mathbb{E}\Big[\mathcal{T}_{1,1}\Big] = \sum_{k=0}^{1}\Big[\mathbb{E}\big[\mathcal{T}_{1,1} \mid \gamma_{1,1} = k\big]P(\gamma_{1,1} = k)\Big] \tag{25}$$

1.2 Posterior variance

The quantity we want to calculate is:

$$V(\alpha_b \mid \boldsymbol{y},\boldsymbol{g}) = V(\sum_{s,l} w_{(s,l),b}\beta_{sl} \mid \boldsymbol{y},\boldsymbol{g}) \tag{26}$$

which cannot be calculated as the quantities are dependent, and is therefore not straightforward to add. To simplify notation, we drop the conditioning on $\boldsymbol{y},\boldsymbol{g}$. The algorithm is as follows:

0. **Initialise:** At the finest scale, $j = J$ (bottom of tree), for state $m = 0, 1$, evaluate, at all locations:

$$V(\beta_{J,l} \mid \gamma_{J,l} = m) \tag{27}$$

Note that in practice, this will be 0 when $m = 0$.

1. At scale j, note that elements at scale $j-1$ are parents of those in scale j. Calculate, for all locations, for all states, $k = 0, 1$:

$$\begin{aligned} &V(\mathcal{T}_{j-1,l} \mid \gamma_{j-1,l} = k) \\ &\qquad = V(E[w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k]) + E[V(w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k)] \end{aligned} \tag{28}$$

For the first term, we have:

$$V(E[w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k]) \tag{29}$$

$$= V[\mu_{j-1,l}(\gamma_{j,c_1}, \gamma_{j,c_2}, k)] \tag{30}$$

$$\begin{aligned} &= \sum_{a=0}^{1}\sum_{m=0}^{1} P(\gamma_{j,c_1} = a \mid \gamma_{j-1,l} = k)P(\gamma_{j,c_2} = m \mid \gamma_{j-1,l} = k)\big[\mu_{j-1,l}(\gamma_{j,c_1} = a, \gamma_{j,c_2} = m, k)^2 + \dots \\ &\sigma^2_{j-1,l}(\gamma_{j,c_1} = a, \gamma_{j,c_2} = m, k)\big]^2 - \dots \\ &\Big[\sum_{a=0}^{1}\sum_{m=0}^{1} P(\gamma_{j,c_1} = a \mid \gamma_{j-1,l} = k)P(\gamma_{j,c_2} = m \mid \gamma_{j-1,l} = k)\big[\mu_{j-1,l}(\gamma_{j,c_1} = a, \gamma_{j,c_2} = m, k)\Big]^2 \end{aligned} \tag{31}$$

For the second term, we have:

$$E[V(w_{j-1,l}\beta_{j-1,l} + \mathcal{T}_{j,c_1} + \mathcal{T}_{j,c_2} \mid \gamma_{j-1,l} = k)] \tag{32}$$

$$E[\sigma^2_{j-1,l}(\gamma_{j,c_1}, \gamma_{j,c_2}, k)] \tag{33}$$

$$\sum_{a=0}^{1}\sum_{m=0}^{1} P(\gamma_{j,c_1} = a \mid \gamma_{j-1,l} = k)P(\gamma_{j,c_2} = m \mid \gamma_{j-1,l} = k)\sigma^2_{j-1,l}(\gamma_{j,c_1} = a, \gamma_{j,c_2} = m, k) \tag{34}$$

Can show that all these terms are known/calculated in previous iterations of the algorithm. Analogously to the expectation quantity last time, due to conditional independence (as we are conditioning on all the states required, and parents, and each subtrees are all independent of each other after conditioning on their states), we have additivity of variances:

$$\sigma^2_{j-1,l}(\gamma_{j,c_1} = a, \gamma_{j,c_2} = m, k) = w^2_{j-1,l}V(\beta_{j-1,l} \mid \gamma_{j-1,l} = k) + V(\mathcal{T}_{j,c_1} \mid \gamma_{j,c_1} = a) + V(\mathcal{T}_{j,c_2} \mid \gamma_{j,c_2} = m) \tag{35}$$

2. Set $j := j - 1$

3. If $j = 1$ (coarsest/least granular level), then stop, else return to step 1.

4. To get the desired quantity, marginalise over all states:

$$V\Big(\mathcal{T}_{1,1}\Big) = \sum_{k=0}^{1}\Big[V\big(\mathcal{T}_{1,1} \mid \gamma_{1,1} = k\big)P(\gamma_{1,1} = k)\Big] \tag{36}$$