Dynamic algorithm for Poisson HMT – posterior means and variances

Brendan Law

August 2, 2019

HJ, you're right – the algorithm works from the bottom to the top of the tree. Here's a quick explanation, hopefully.

1 Notation

Note that, at each scale-location, sl, where $s = 1, \ldots, J$ and $l = 1, \ldots, 2^{J-1}$ of our tree, we actually have two quantities of interest at each node for this calculation. We have $\log p_{s,l}$ and $\log q_{s,l}$ (currently ignoring covariate groups, denoted with subscripts ^{1} and ^{0} in *multiseq*).

Our dynamic algorithm calculates all p, q quantities at all scales and locations simultaneously, hence we could notate as thus:

$$\eta_{s,l}, \quad \text{for } s = 1, \ldots, J, l = 1, \ldots, 2^J$$

Note carefully now that $l = 1, \ldots, 2^J$ as there are two parameters required at each node of the tree.

Because our algorithm works from the bottom to the top of the tree, it's easiest to flip the standard scale convention, and have that $s = 1$ represent the leaves (bottom) of the tree, and $s = J$ is the root (top) of the tree. Therefore, we have:

$$\begin{aligned}
\eta_{J,1} &:= \log p_{1,1} \\
\eta_{J,2} &:= \log q_{1,1} \\
&\vdots \\
\eta_{1,1} &:= \log p_{J,1} \\
\eta_{1,2^J} &:= \log q_{J,2^{J-1}}
\end{aligned}$$

The data we have observed are the MLE estimates of the parameters at each scale-location:

$$\begin{aligned}
d_{s,l} &= (\widehat{\beta}_{s,l}, \text{se}(\widehat{\beta}_{s,l})^2, \widehat{\mu}^*_{s,l}, \text{se}(\widehat{\mu}^*_{s,l})^2) \\
D_j &= (d_{s,l})_{s=1}^{j}
\end{aligned}$$

with D_j being a collection of all data at all nodes from the bottom of the tree all the way up until scale j of the tree.

From previous work, we know, for each sl, the conditional means and variances, for each of the $m = 0, \ldots, M$ states in our mixture prior distribution (the zero-th state is our point mass at zero):

$$\begin{aligned}
E[\eta_{s,l} \mid D_s, \gamma_{s,l} = m] \\
V[\eta_{s,l} \mid D_s, \gamma_{s,l} = m]
\end{aligned}$$

We also know, as output from our EM algorithm:

$$\begin{aligned}
P(\gamma_{s,l} \mid D_J) \\
P(\gamma_{s,l}, \gamma_{p(s,l)} \mid D_J) \\
\therefore P(\gamma_{s,l} \mid \gamma_{p(s,l)}, D_J)
\end{aligned}$$

with $p(s, l)$ being the parent of s, l.

We also have that:

$$E[X] = E[E[X \mid Y]] \tag{1}$$
$$V(X) = V(E[X \mid Y]) + E[V(X \mid Y)] \tag{2}$$

2 Objectives

Our end goal is calculate two quantities. For each 'sequence' of p's and q's, $(\eta_1, \ldots, \eta_J)$, required to calculate a particular data-space quantity at a base b, we want the posterior expectations and variances:

$$E[\sum_{s=1}^{J} \eta_J \mid D_J]$$
$$V[\sum_{s=1}^{J} \eta_J \mid D_J]$$

3 Algorithm

Idea is to use *dynamic programming*; start at the smallest instance, and then calculate larger instances as a collection of the (previously-calculated) smaller instances. Hence, we start at the bottom of the tree, and work our way up.

Initialise:
Set $s = 1$; the finest scale (bottom of tree). Across all locations, $l = 1, \ldots, 2^J$, and across all states, $m = 0, \ldots, M$, calculate:

$$E[\eta_{1,l} \mid \gamma_{1,l} = m, D_1]$$

Step 1:
At $s = j$, for all states, $m = 0, \ldots, M$, and for all locations, l:

$$E[\eta_{p(j,l)} + \sum_{s=1}^{j} \eta_{s,l} \mid \gamma_{p(j,l)} = m, D_{j+1}] =$$
$$\sum_{n=0}^{M} E\Big[\eta_{p(j,l)} + \sum_{s=1}^{j} \eta_{s,l} \mid \gamma_{p(j,l)} = m, \gamma_{j,l} = n, D_{j+1}\Big] P\Big(\gamma_{j,l} = n \mid \gamma_{p(j,l)} = m, D_{j+1}\Big) \qquad \text{(by (1))}$$

Where the probability comes out of the HMT EM algorithm, and the expectation can be calculated:

$$E\Big[\eta_{p(j,l)} + \sum_{s=1}^{j} \eta_{s,l} \mid \gamma_{p(j,l)} = m, \gamma_{j,l} = n, D_{j+1}\Big] =$$
$$E\Big[\eta_{p(j,l)} \mid \gamma_{p(j,l)} = m, d_{p(j,l)}\Big] + E\Big[\sum_{s=1}^{j} \eta_{s,l} \mid \gamma_{j,l} = n, D_j\Big] \tag{3}$$

The first quantity of (3), we get from previous calculations, and the second quantity of (3), we will calculate dynamically through repeated iterations of step 1.

Step 2:
Set $j := j + 1$.

Step 3:
If $j = J$ (root of tree), then stop, else return to step 1.

Step 4:
Up until this time, we have calculated, for all locations (all p, q 'sequences' required for a data-space quantity):

$$E\Big[\sum_{s=1}^{J} \eta_{s,l} \mid \gamma_{J,l} = m, D_J\Big]$$

Hence, we get the posterior mean by marginalising, using probabilities of the state of the root, obtained from our EM HMT algorithm:

$$E\Big[\sum_{s=1}^{J}\eta_{s,l} \mid D_J\Big] = \sum_{m=0}^{M} E\Big[\sum_{s=1}^{J}\eta_{s,l} \mid \gamma_{J,l}=m, D_J\Big] P\Big(\gamma_{J,l}=m \mid D_J\Big)$$

For variances, a similar algorithm applies – we can swap all the expectation operations for variance operations. We can do this, as the variances are additive in this situation:

- We have conditional independence of each of the pair of quantities we consider at each step (as a coefficient of the tree is independent of all other random variables, given its own state – i'm pretty sure the Poisson decomposition retains this wavelet-tree-like property?). η are functions of p and q, which are functions of α, hence all independent?
- Also, our conditional variance formula in step 1 would be:

$$\begin{aligned} V[\eta_{p(j,l)} + \sum_{s=1}^{j}\eta_{s,l} \mid \gamma_{p(j,l)}=m, D_{j+1}] = \\ \sum_{n=0}^{M} V\Big[\eta_{p(j,l)} + \sum_{s=1}^{j}\eta_{s,l} \mid \gamma_{p(j,l)}=m, \gamma_{j,l}=n, D_{j+1}\Big] P\Big(\gamma_{j,l}=n \mid \gamma_{p(j,l)}=m, D_{j+1}\Big) + \\ V\Big(E\Big[\eta_{p(j,l)} + \sum_{s=1}^{j}\eta_{s,l} \mid \gamma_{p(j,l)}=m, \gamma_{j,l}=n, D_{j+1}\Big]\Big) \end{aligned}$$

  but the second term of the sum equals 0, as this expectation is a deterministic quantity.