Structured Variational Approximation for Gaussian Assembly Graphs

Christopher Quince

1 The model

Define the joint distribution of:

- Transformed counts $x_{v,s} = \sqrt{x}'_{v,s}$ for each unitig $v = 1, \dots, V$ in sample $s = 1, \dots, S$
- Paths for strain $g = 1, \dots, G$ defined by $\eta^g_{u,v}$
- Flow of strain g through v, $\phi^{g+}_v = \sum_{u \in A(v)} \eta^g_{u,v}$ and $\phi^{g-}_v = \sum_{u \in D(v)} \eta^g_{v,u}$ where $A(v)$ is set of ancestors of v and $D(v)$ descendants in the assembly graph
- Strain coverages $\gamma_{g,s}$
- Unitig lengths L_v
- Source node s and sink node t

$$P(\mathbf{X}, \mathbf{\Gamma}, \mathbf{H}) = \prod_{v=1}^{V} \prod_{s=1}^{S} \mathcal{N}(x_{v,s} | L_v [\sum_{h=1}^{G} \phi^h_v \gamma_{h,s}], \sigma^2) \prod_{h=1}^{G} \prod_{s=1}^{S} P(\gamma_{g,s}) \prod_{h=1}^{G} \prod_{v=1}^{V} \delta_{\phi^{g+}_v, \phi^{g-}_v} \delta_{\phi^{g-}_s, 1} \delta_{\phi^{g+}_t, 1} \tag{1}$$

where $\sigma^2 = 1/4$ and δ is the Kronecker delta.

We assume an exponential prior for the $\gamma_{g,s}$ such that:

$$P(\gamma_{g,s}) = \frac{1/\epsilon}{\exp}(-\gamma_{g,s}/\epsilon) \tag{2}$$

2 Variational Approximation

Assume the following factorisation for the variational approximation:

$$q(\mathbf{X}, \mathbf{\Gamma}, \mathbf{H}) = \prod_{h=1}^{G} q_h(\{\eta^h_{v,u}\}_{u,v \in A}) \prod_{h=1}^{G} \prod_{s=1}^{S} q_g(\gamma_{h,s}) \tag{3}$$

where A are all pairs of nodes in assembly graph.

Then the mean field update for each set of $\{\eta^h_{v,u}\}_{u,v\in A}$ is derived as:

$$
\begin{aligned}
\ln q^*(\{\eta^h_{v,u}\}_{u,v\in A}) &= \langle \ln P \rangle_{\phi^{h\neq g}_{v\in A},\gamma_{g,s}} && (4)\\
&= \ln\left(\prod_{v=1}^{V} \delta_{\phi^{g+}_v,\phi^{g-}_v}\delta_{\phi^{g-}_s,1}\delta_{\phi^{g+}_t,1}\right) && (5)\\
&- \left\langle \sum_{v=1}^{V}\sum_{s=1}^{S}\frac{1}{2\sigma^2}\left(x_{v,s} - L_v[\sum_{h=1}^{G}\phi^h_v\gamma_{h,s}]\right)^2 \right\rangle_{\phi^{h\neq g}_{v\in A},\gamma_{g,s}} && (6)
\end{aligned}
$$

Consider second term only:

$$
-\frac{1}{2\sigma^2}\left(-\sum_{v=1}^{V}\sum_{s=1}^{S}2x_{v,s}L_v\langle\gamma_{g,s}\rangle\phi^g_v + L_v^2\langle(\sum_{h=1}^{G}\phi^h_v\gamma_{h,s})(\sum_{g=1}^{G}\phi^g_v\gamma_{g,s})\rangle)\right) \tag{7}
$$

$$
-\frac{1}{2\sigma^2}\left(\sum_{v=1}^{V}\sum_{s=1}^{S}\left[-2x_{v,s}L_v\langle\gamma_{g,s}\rangle\phi^g_v + 2L_v^2\sum_{h\neq g}^{G}\langle\phi^h_v\rangle\langle\gamma_{h,s}\rangle\langle\gamma_{g,s}\rangle\phi^g_v + L_v^2\langle\gamma^2_{g,s}\rangle(\phi^g_v)^2\right]\right) \tag{8}
$$