0.1 Combining modalities and data

Note that the matrix $\mathbf{B}$ – and also the matrix $\mathbf{C}$ – can get very large very quickly: The former is of size $2^V \times 2^{V \cdot |\mathcal{O}|}$ and the latter has dimensions $2^{V \cdot |\mathcal{O}|} \times N$, meaning both grow exponentially with the number of lymph node levels (LNLs) *and* diagnostic modalities. And although neither $\mathbf{B}$ not $\mathbf{C}$ depend on the parameters θ, meaning their product can be precomputed, we can simply iterate over all patients, possible hidden states and available diagnostic modalities to compute $\mathbf{\Omega} \coloneqq \mathbf{B} \cdot \mathbf{C}$ directly, which saves us building up and multiplying matrices with potentially millions of entries.

To compute this matrix $\mathbf{\Omega}$, we first abandon the just-introduced way of combining diagnoses for all modalities into one large vector and separate them again, so that we have complete and incomplete observations $\boldsymbol{\zeta}_j^k$ and $\mathbf{d}_n^k$ respectively for each modality, where $n \in [1, N]$ enumerates the patients in the data.

$$\begin{aligned} \Omega_{mn} = P\left(\mathbf{d}_n \mid \boldsymbol{\xi}_m\right) &= \prod_{k=1}^{|\mathcal{O}|} P\left(\mathbf{d}_n^k \mid \boldsymbol{\xi}_m\right) \\ &= \prod_{k=1}^{|\mathcal{O}|} \left[\sum_{j \,:\, \mathrm{match}(\mathbf{d}_n^k, \boldsymbol{\zeta}_j^k)} P\left(\boldsymbol{\zeta}_j \mid \boldsymbol{\xi}_m\right) \right] = \prod_{k=1}^{|\mathcal{O}|} \left[\sum_{j \,:\, \mathrm{match}(\mathbf{d}_n^k, \boldsymbol{\zeta}_j^k)} B_{mj}^k \right] \end{aligned} \tag{1}$$

Now, the elements Ω_{mn} encode the observation likelihood of patient n's diagnose $\mathbf{d}_n$ given their true state of involvement is $\boldsymbol{\xi}_m$. Finally, with this the row-vector of likelihoods of a cohort of patients, given the model's spread parameters, becomes

$$P\left(\boldsymbol{\mathcal{D}} \mid \theta\right) = p_T\left(\mathbf{t}\right) \cdot \mathbf{\Lambda} \cdot \mathbf{\Omega} \tag{2}$$

Again, the objects $p_T(\mathbf{t})$ and $\mathbf{\Lambda}$ depend on the parameters and hence need to be recalculated for every sample drawn during Markov-chain Monte Carlo (MCMC) inference. $\mathbf{\Omega}$ depends only on the patient data $\mathcal{D}$ and must therefore only be computed once at the beginning of the learning round.