0.1 Inference and risk assessment for incomplete diagnoses

A diagnosis is often not complete, meaning that not all lymph node levels (LNLs) might have been checked with a diagnostic modality. E.g., fine needle aspiration (FNA) is usually only performed in a subset of LNLs. Hence, we must be able to deal with "incomplete" observations for some LNLs. To do so, we first introduce a new observation variable

$$d_v \in \{0, 1, \emptyset\} \tag{1}$$

where $\emptyset$ indicates *unobserved*. Furthermore, we define a *match function*

$$\text{match}(\mathbf{d}, \mathbf{z}) \coloneqq \begin{cases} \text{true} & \text{if } d_v = z_v \vee d_v = \emptyset;\ \forall v \\ \text{false} & \text{else} \end{cases} \tag{2}$$

which returns *true* if a - potentially incomplete - diagnosis $\mathbf{d}$ is consistent with a complete observation $\mathbf{z}$. We will use this function for conveniently marginalizing over the missing observations. In analogy to **??**, we can compute the risk for an incomplete observation as

$$\begin{aligned} R\left(X_v = 1 \mid \mathbf{d}, \theta\right) &= \frac{P\left(\mathbf{d} \mid X_v = 1, \theta\right) P\left(X_v = 1 \mid \theta\right)}{P\left(\mathbf{d} \mid \theta\right)} \\ &= \sum_{i\,:\,\xi_{iv}=1} \frac{P\left(\mathbf{d} \mid \boldsymbol{\xi}_i, \theta\right) P\left(\boldsymbol{\xi}_i \mid \theta\right)}{P\left(\mathbf{d} \mid \theta\right)} \end{aligned} \tag{3}$$

where the enumerator of the second line can now be rewritten using the match function:

$$\begin{aligned} P\left(\mathbf{d} \mid \boldsymbol{\xi}_i, \theta\right) P\left(\boldsymbol{\xi}_i \mid \theta\right) &= \sum_{\left\{j\,:\,\text{match}(\mathbf{d}, \boldsymbol{\zeta}_j)\right\}} P\left(\boldsymbol{\zeta}_j \mid \boldsymbol{\xi}_i, \theta\right) P\Big(\boldsymbol{\xi}_i \mid \theta\Big) \\ &= \sum_{\left\{j\,:\,\text{match}(\mathbf{d}, \boldsymbol{\zeta}_j)\right\}} B_{ij} \Big[p_T\left(\mathbf{t}\right) \cdot \boldsymbol{\Lambda}\Big]_i \end{aligned} \tag{4}$$

In this case B_{ij} denotes the element of the observation matrix that corresponds to state $\boldsymbol{\xi}_i$ and observation $\boldsymbol{\zeta}_j$. Again, the indices $\{i\ :\ \xi_{iv} = 1\}$ in eq. (3) correspond to all possible states with a positive involvement in lymph node level X_v. Essentially, the whole term is the likelihood of an observation $\mathbf{d}$ where we have removed all entries that correspond to states with $X_v \neq 1$ both from the observation matrix and the resulting probability vector of the evolution. It can therefore be easily computed algebraically, too.

The evidence in the denominator of eq. (3) becomes a marginalization over all possible diagnoses that are not available to us or that we deem unimportant

$$P\left(\mathbf{d} \mid \theta\right) = \sum_{\left\{j\,:\,\text{match}(\mathbf{d}, \boldsymbol{\zeta}_j)\right\}} \Big[p_T\left(\mathbf{t}\right) \cdot \boldsymbol{\Lambda}\Big]_j \tag{5}$$

We can make this summation a bit more elegant using a column vector $\mathbf{c}^{\mathbf{d}}$ that has entries corresponding to the match-function

$$c_i^{\mathbf{d}} = \text{match}(\mathbf{d}, \boldsymbol{\zeta}_i) \tag{6}$$

where every *true* corresponds to a 1 and every *false* to a 0. This way we can rewrite eq. (5) in the following way:

$$P\left(\mathbf{d} \mid \theta\right) = p_T\left(\mathbf{t}\right) \cdot \mathbf{\Lambda} \cdot \mathbf{B} \cdot \mathbf{c}^{\mathbf{d}} \tag{7}$$

Using this notation for marginalizing over unknown or incomplete observations also allows us to encode entire datasets $\boldsymbol{\mathcal{D}} = \begin{pmatrix}\mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_N\end{pmatrix}$ of (potentially incomplete) observations in the form of a matrix

$$\mathbf{C} = \begin{pmatrix}\mathbf{c}^{\mathbf{d}_1} & \mathbf{c}^{\mathbf{d}_2} & \cdots & \mathbf{c}^{\mathbf{d}_N}\end{pmatrix} \tag{8}$$

so that the row-vector of likelihoods reads as

$$P\left(\boldsymbol{\mathcal{D}} \mid \theta\right) = \left(P\left(\mathbf{d}_n \mid \theta\right)\right)_{n\in[1,N]} = p_T\left(\mathbf{t}\right) \cdot \mathbf{\Lambda} \cdot \mathbf{B} \cdot \mathbf{C} \tag{9}$$