0.1 Incorporation of T-stage

We have introduced the hidden Markov model (HMM) with the promise that it could handle the concept of T-stages through its explicit modeling of dynamic processes. To keep up with that, we will now explain how this is achieved using the time-prior $p(t)$.

The core idea is to assume that early T-stage and late T-stage tumors share the same patterns of metastatic progression, except that late T-stage tumors are on average diagnosed at a later point in time, and thereby also show, on average, higher lymph node level (LNL) involvement. Formally, this can be described by assuming a different time-prior $p_T(t)$ for every T-stage T. On the other hand, the transition matrix $\mathbf{A}$ is assumed to be the same for all T-stages.

For the inference of model parameters, the training data is split into subgroups according to T-stage. We now define a column-vector $\mathbf{f}_T$ separately for each T-stage, which counts the number of patients in the dataset that were diagnosed with one of the possible observational states and a given T-stage. The log-likelihood from which we want to sample is then simply a sum of the likelihoods as above, where the essential difference is that we equip each marginalization over time with a different time-prior $p_T(t)$, according to its T-stage:

$$\log P\left(\boldsymbol{\mathcal{Z}} \mid \theta\right) = \sum_{T=1}^{4} \log \left[\sum_{t=0}^{t_{\max}} p_T(t) \cdot \boldsymbol{\pi}^\top \cdot (\mathbf{A})^t \cdot \mathbf{B} \right] \cdot \mathbf{f}_T \tag{1}$$

The logarithm must be taken element-wise for the resulting row-vector inside the square brackets. The only data-dependent term here is the vector $\mathbf{f}_T$ counting the occurrences of all possible observations. It is again important to note that the only difference between the part of the log-likelihood for the different T-stages is the exact shape or parametrization of the time-prior. The transition probabilities, and hence also the transition matrix $\mathbf{A}$, are the same for all T-stages. For this to work, we rely on the assumption that different typical patterns of nodal involvement for the same primary tumor location are caused mainly by different progression times

At this point, it makes sense to briefly introduce a notation of the above equation that is more suitable for the actual programmatic implementation of the inference and the extension we will discuss later. We can rewrite the term in the square brackets of eq. (1) by using the matrix

$$\mathbf{\Lambda} \coloneqq P\left(\mathbf{X} \mid \mathbf{t}\right) = \begin{pmatrix} \boldsymbol{\pi}^\top \cdot (\mathbf{A})^0 \\ \boldsymbol{\pi}^\top \cdot (\mathbf{A})^1 \\ \vdots \\ \boldsymbol{\pi}^\top \cdot (\mathbf{A})^{t_{\max}} \end{pmatrix} \tag{2}$$

were row number t corresponds to the vector $\boldsymbol{\pi}^\top \cdot (\mathbf{A})^t$, i.e. the probabilities for all possible hidden states, given the diagnose time. So, the element $\mathbf{\Lambda}_{ti}$ corresponds to the probability $P\left(\boldsymbol{\xi}_i \mid t\right)$ of a patient arriving in the ith state after t time steps. With this, we can rewrite the term in the square brackets of eq. (1) purely as a product of vectors and matrices:

$$\sum_{t=0}^{t_{\max}} p_T(t) \cdot \boldsymbol{\pi}^\top \cdot (\mathbf{A})^t = p_T\left(\mathbf{t}\right) \cdot \mathbf{\Lambda} \tag{3}$$

with $p_T(\mathbf{t}) = \begin{pmatrix} p_T(0) & p_T(1) & \cdots & p_T(t_{\max}) \end{pmatrix}$. The matrix $\mathbf{\Lambda}$ implicitly depends on the spread probabilities, while each of the $p_T(\mathbf{t})$ depends on the respective parametrization of the time prior. They are the only objects that depend on the parameters θ and they are independent of the data.