0.1 Marginalization

To calculate the likelihood function, we have to calculate the probability of a given diagnostic observation. To that end, we first calculate the probability of observing a given diagnosis $\mathbf{z} = \boldsymbol{\zeta}_j$ at a fixed time-step t. As depicted in **??**, we must consider every possible evolution of a patient's disease that leads to the observed diagnosis. Mathematically, this means that we need to marginalize over all such paths. And here is where the HMM-formalism comes in very useful, because this marginalization happens automatically when we multiply the transition matrix with itself and eventually with the observation matrix:

$$P\left(\mathbf{Z} = \boldsymbol{\zeta}_j \mid t\right) = \left[\boldsymbol{\pi}^\top \cdot (\mathbf{A})^t \cdot \mathbf{B}\right]_j \tag{1}$$

where the $\boldsymbol{\pi}$ is the column vector for the healthy starting state. $\mathbf{A}$ is multiplied with itself t times and thereby produces a matrix that describes the transition probability from the healthy state to all possible states $\mathbf{x}[t]$ in exactly t time-steps marginalized over the actual pathway of the patient's disease. The index $[\ldots]_j$ here means that from the resulting (row-)vector of probabilities we take the component that corresponds to the diagnose $\mathbf{z} = \boldsymbol{\zeta}_j$.

So, essentially, eq. (1) first computes the probability vector of all possible true hidden states, given a time step t

$$P\left(\mathbf{X} = \boldsymbol{\xi}_i \mid t\right) = \left[\boldsymbol{\pi}^\top \cdot (\mathbf{A})^t\right]_i \tag{2}$$

and then multiplies it with the respective observation probability vector, which is a column of the $\mathbf{B}$ matrix, to finally marginalize over all possible true hidden states – effectively a sum over i in eq. (2) – at the time t of diagnosis.

The problem that the number of time-steps until diagnosis is unknown cannot be solved in such an elegant fashion. Therefore, we must resort to brute force marginalization and introduce a prior $p(t)$, which is a discrete distribution over a finite number of time-steps. It describes the prior probability that a patient's cancer is diagnosed at a particular time-step t. To get the probability of a diagnosis $\mathbf{z}$ we must compute

$$P\left(\mathbf{Z} = \boldsymbol{\zeta}_j\right) = \sum_{t=0}^{t_\text{max}} p(t) \cdot P\left(\mathbf{Z} = \boldsymbol{\zeta}_j \mid t\right) = \left[\sum_{t=0}^{t_\text{max}} p(t) \cdot \boldsymbol{\pi}^\top \cdot (\mathbf{A})^t \cdot \mathbf{B}\right]_j \tag{3}$$

While the choice of the time-prior may seem unclear at this point, its role for including T-stage into this model will be discussed in **??**.