0.1 Risk assessment of microscopic involvement

With a parameter set $\theta = \left(\left\{\tilde{b}_v\right\}, \left\{\tilde{t}_{rv}\right\}_{r \in \mathrm{pa}(v)}\right) \; \forall v \leq V$, we can assess the risk of nodal involvement, given a diagnosis $\mathbf{z}$, of a new patient. Using Bayes' law, the risk for a certain lymph node level (LNL) v being involved is given by the conditional probability

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

Note that in the second line, we have explicitly written out the marginalization over all hidden states $\boldsymbol{\xi}_i$ that have LNL v involved. We have written the state of LNL v in the state $\boldsymbol{\xi}_i$ as ξ_{iv}. The denominator can be computed using **??**, which already includes the marginalization over all hidden states $\boldsymbol{\xi}_i$.

The process of sampling randomly generates L sets of parameters $\theta = \begin{pmatrix}\theta_1 & \theta_2 & \dots & \theta_L\end{pmatrix}$. They are therefore random variables and so is the risk $R\left(X_v \mid \mathbf{z}, \theta\right)$ since it is a function of θ. Using the Monte Carlo estimator, we can therefore compute the moments of the distribution over the risk, including e.g. the expectation value

$$\mathbb{E}_{\boldsymbol{\theta}}\left[R\left(X_v = 1 \mid \mathbf{z}\right)\right] = \frac{1}{L} \sum_{k=1}^{L} R\left(X_v = 1 \mid \mathbf{z}, \theta_k\right) \tag{2}$$

In the result sections below, we compute the individual risks for a large enough number L of sampled parameters. Thereby, we can compute histograms for the risk that will approach the real probability density of the respective risk for $L \to \infty$. This provides additional information on the uncertainty in the predicted risk resulting from uncertainty in the model parameters.