

---

## 0.1 Sampling

With a parameter set  $\theta = \left( \{\tilde{b}_v\}, \{\tilde{t}_{rv}\}_{r \in \text{pa}(v)} \right) \forall v \leq V$ , 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

$$\begin{aligned} R(X_v = 1 \mid \mathbf{z}, \theta) &= \frac{P(\mathbf{Z} = \mathbf{z} \mid X_v = 1, \theta) P(X_v = 1 \mid \theta)}{P(\mathbf{Z} = \mathbf{z} \mid \theta)} \\ &= \sum_{i: \xi_{iv}=1} \frac{P(\mathbf{Z} = \mathbf{z} \mid \xi_i, \theta) P(\xi_i \mid \theta)}{P(\mathbf{Z} = \mathbf{z} \mid \theta)} \end{aligned} \quad (1)$$

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

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

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

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  $L \rightarrow \infty$ . This provides additional information on the uncertainty in the predicted risk resulting from uncertainty in the model parameters.