

# Chapter 1

## Bilateral hidden Markov model

In the previous chapter we have set up the formalism to deal with only one side of the neck. Implicitly, we have assumed that to be the ipsilateral side, i.e. the side of the sagittal plane where the primary tumor is located. This is because we assume lymphatic drainage to a process that is somewhat symmetric w.r.t. the sagittal plane, which means there can only be limited lymph flow across this plane. But depending on the tumor's location and lateralization, drainage and hence metastatic spread to the contralateral lymphatic system of the neck may also occur. In current clinical practice, a bilateral neck dissection or irradiation is often prescribed when the tumor is close to the mid-sagittal plane. So, ideally, we would like to model the risk for involvement in both sides of the neck at the same time.

The formalism of ?? can easily be applied to the contralateral side and given respective training data for the sampling process, it would learn the appropriate spread probabilities to and among the contralateral lymph node levels (LNLs) just as it would learn the ones for the ipsilateral side. From clinical experience, the contralateral involvement is usually less severe than the ipsilateral one, and hence we would expect the contralateral spread to be less probable as well.

However, combining two such unilateral models naively would make the assumption that ipsi- and contralateral spread are independent, which seems unlikely: If we know a patient has advanced metastases in the contralateral neck nodes, the risk to find similarly or even more advanced disease in ipsilateral neck nodes should probably be higher than if the contralateral neck were healthy. In other words, we are now looking for the joint probability  $P(\mathbf{X}^i, \mathbf{X}^c | \mathbf{Z}^i, \mathbf{Z}^c)$ , where the superscripts i and c indicate the ipsi- and contralateral side respectively.

The following section will pick up the unilateral formalism, extend and modify it to come up with a less naive bilateral model.

### 1.1 Expanding the unilateral model

If we start by dissecting this joint conditional probability in the following way

$$P(\mathbf{X}^i, \mathbf{X}^c | \mathbf{Z}^i, \mathbf{Z}^c) = \frac{P(\mathbf{Z}^i, \mathbf{Z}^c | \mathbf{X}^i, \mathbf{X}^c) \cdot P(\mathbf{X}^i, \mathbf{X}^c)}{P(\mathbf{Z}^i, \mathbf{Z}^c)} \quad (1.1)$$

we notice right away that the likelihood on the right factorizes: Given the true states of involvement in the two sides of the neck, their respective diagnoses must

be independent. Furthermore, the two factors are already given by their corresponding observation matrices  $\mathbf{B}^i$  and  $\mathbf{B}^c$ .

The joint probability of the hidden states  $P(\mathbf{X}^i, \mathbf{X}^c)$  does not factorize in the same manner. But if we assume the lymphatic network to be symmetric and directed, there can be no direct connection between LNLs of the two sides of the neck, which means the probability for involvement of the ipsi- and contralateral side only correlate via the diagnose time  $t$ . Hence the joint probability is a sum of factorizing terms:

$$\begin{aligned} P(\mathbf{X}^i, \mathbf{X}^c) &= \sum_{t \in \mathbb{T}} p(t) \cdot P(\mathbf{X}^i, \mathbf{X}^c | t) \\ &= \sum_{t \in \mathbb{T}} p(t) \cdot P(\mathbf{X}^c | t) \cdot P^\top(\mathbf{X}^i | t) \end{aligned} \quad (1.2)$$

Note that the two row vectors of probabilities in the second line are multiplied using an outer product. Using the notation from the last section, We can write this in an algebraic way to effectively factorize this sum as follows

$$P(\mathbf{X}^c = \boldsymbol{\xi}_n, \mathbf{X}^i = \boldsymbol{\xi}_m) = [\boldsymbol{\Lambda}_c^\top \cdot \text{diag } p(\mathbf{t}) \cdot \boldsymbol{\Lambda}_i]_{n,m} \quad (1.3)$$

where the  $\boldsymbol{\Lambda}$  are again matrices with rows of the conditional probabilities  $P(\mathbf{X} | t)$  which can be computed as defined in ???. Multiplying these two matrices – one for the contralateral side from the left and one for the ipsilateral side from the right – onto a diagonal matrix containing the time prior marginalizes over the diagnose time and results in a matrix where the value in row  $n$  and column  $m$  represents the probability to find the contralateral neck in state  $\mathbf{X}^c = \boldsymbol{\xi}_n$  and the ipsilateral lymphatic system in state  $\mathbf{X}^i = \boldsymbol{\xi}_m$ .

Similarly, we can now multiply the observation matrices  $\mathbf{B}$  from the left and the right onto  $P(\mathbf{X}^i, \mathbf{X}^c)$  to compute the bilateral equivalent of ??:

$$P(\mathbf{Z}^c = \boldsymbol{\zeta}_n, \mathbf{Z}^i = \boldsymbol{\zeta}_m) = [\mathbf{B}^\top \cdot \boldsymbol{\Lambda}_c^\top \cdot \text{diag } p(\mathbf{t}) \cdot \boldsymbol{\Lambda}_i \cdot \mathbf{B}]_{n,m} \quad (1.4)$$

Formally, all necessary terms can now be computed so that both inference and the subsequent risk prediction can be performed. However, in the next section we will go into more detail regarding how this was implemented.

## 1.2 Parameter symmetries and mid-line extension

Although it has been omitted, eqs. (1.1) to (1.4) are still functions of the same parameters as in the unilateral model, but each side now has their own set  $\boldsymbol{\theta}^c$  and  $\boldsymbol{\theta}^i$  of spread probabilities that are used to parameterize the transition matrices  $\mathbf{A}^c$  and  $\mathbf{A}^i$  respectively.

In principle, the spread probabilities of the two sides are entirely indepent, and a lateralized primary tumor certainly spreads to a different extend to the ipsi- versus the contralateral side. But the spread probabilities among the LNLs should be equal when assuming that the lymphatic network in the head and neck region is symmetric. This means

$$\begin{aligned} \tilde{b}_v^c &\neq \tilde{b}_v^i \\ \tilde{t}_{rv}^c &= \tilde{t}_{rv}^i \end{aligned} \quad \forall v \leq V, r \in \text{pa}(v) \quad (1.5)$$

Due to the reasonable assumption of a symmetric neck anatomy, we may avoid doubling the spread parameters when we model the bilateral lymphatic spread.

However, there are cases in which the primary tumor lies almost or exactly on the mid-sagittal plane of the patient. In such cases, we cannot reasonably distinguish between the ipsi- and contralateral side. Consequently, we must assume the base probability rates as well to be symmetric:  $\tilde{b}_v^c = \tilde{b}_v^i$

This means there must be a continuous increase in the spread probabilities from the primary tumor to the contralateral LNLs if we were to move a patient's tumor from a clearly lateralized location closer and closer to that patient's mid-sagittal plane. Ideally, we would like to factor information about the tumor's "degree of asymmetry" into our model, e.g. by considering a normalized perpendicular distance from the mid-sagittal plane to the tumor's center of mass or by considering the tumor volume on either side of this plane. Data like this, however, is rarely available. What is frequently reported and also clinically considered as a risk factor for contralateral involvement is whether or not the tumor touches or extends over the mid-sagittal plane. With this binary variable (and the information on whether the tumor is central/symmetric w.r.t. to the sagittal symmetry plane) we can now distinguish three degrees of lateralizations:

1.  $\not{s}, \not{e}$ : The tumor does not cross or touch the mid-sagittal plane and is thus clearly lateralized. The base spread probabilities are  $\{\tilde{b}_v^i\}$  and  $\{\tilde{b}_v^{c,\not{e}}\}$ .
2.  $\not{s}, e$ : The tumor is lateralized, but crosses or touches the mid-sagittal plane. We will discuss how to define the spread probabilities to the contralateral side below.
3.  $s, e$ : The tumor is symmetric w.r.t. to the sagittal plane, thus  $\tilde{b}_v^{c,s} = \tilde{b}_v^i$

Note that  $s(\not{s})$  and  $e(\not{e})$  denote the two binary variables *symmetric* (or *not symmetric*) and *extending* (or *not extending*) over the mid-sagittal plane.

We can infer that in case 2 the spread probabilities to the contralateral LNLs must be between the ones for the clearly lateralized (1) and the symmetric (3) case. Hence, we introduce a new "mixing" parameter  $\alpha$  that defines the contralateral spread from tumor to the LNLs as a linear superposition between the two extremes:

$$\tilde{b}_v^{c,e} = \alpha \cdot \tilde{b}_v^i + (1 - \alpha) \cdot \tilde{b}_v^{c,\not{e}} \quad (1.6)$$

This new mixing parameter must be inferred from data just like the other spread probabilities and the parametrization of the time prior.

When using the learned parameters to predict the risk of a new patient  $g$ , the set of parameters for the risk computation  $\hat{\theta}_g$  is compiled from the total set of inferred parameters  $\hat{\theta} = \{\tilde{b}_v^i, \tilde{b}_v^{c,\not{e}}, \alpha, \tilde{t}_{rv}, p_T\}$ , depending on the risk factors the patient presents with at the time of diagnosis. As always, for  $\hat{\theta}$  we have  $v \leq V$ ,  $r \in \text{pa}(v)$  and the T-stage  $T \in \{1, 2, 3, 4\}$ . For example, if patient  $g$  has a T1 tumor that is clearly lateralized, their  $\hat{\theta}_g$  may be computed in the following way:

$$\hat{\theta}_g = \left\{ \tilde{b}_v^i, \tilde{b}_v^c = \tilde{b}_v^{c,\not{e}}, \tilde{t}_{rv}, p_1 \right\} \quad (1.7)$$

while another patient  $m$  with a T3 tumor that clearly crosses the mid-sagittal plane would have the following set of parameters used for their risk prediction:

$$\hat{\theta}_m = \left\{ \tilde{b}_v^i, \tilde{b}_v^c = \alpha \cdot \tilde{b}_v^i + (1 - \alpha) \cdot \tilde{b}_v^{c,\not{e}}, \tilde{t}_{rv}, p_3 \right\} \quad (1.8)$$

In the actual computational implementation of this model, we essentially compute three different matrices  $\mathbf{\Lambda}$  which are functions of different parameters:

$$\begin{aligned}\mathbf{\Lambda}_i &= \mathbf{\Lambda} \left( \tilde{b}_v^i, \tilde{t}_{rv} \right) \\ \mathbf{\Lambda}_{c,\ell} &= \mathbf{\Lambda} \left( \tilde{b}_v^{c,\ell}, \tilde{t}_{rv} \right) \\ \mathbf{\Lambda}_{c,e} &= \mathbf{\Lambda} \left( \alpha, \tilde{b}_v^{c,\ell}, \tilde{b}_v^i, \tilde{t}_{rv} \right)\end{aligned}\tag{1.9}$$

From those, the likelihoods of all patients in the training data can be computed when used with the respective  $p_T$  – that gives rise to the corresponding  $\text{diag } p(\mathbf{t})$  – as in eq. (1.4).