

Lecture 7

Lecturer: Daniel A. Spielman

To begin, let me point out that there was a typo in the lecture notes from last lecture. Lemma 6.4.1 should have said:

Lemma 7.0.1.

$$P^{ext} [X_1 = a_1 | Y_2 Y_3 = b_2 b_3] = \sum_{a_2: (a_1, a_2) \in \mathcal{C}_{12}} P [X_2 = a_2 | X_1 = a_1] P^{ext} [X_2 = a_2 | Y_2 = b_2] P^{ext} [X_2 = a_2 | Y_3 = b_3].$$

7.1 Markov Property

Last lecture, we considered three variables X_1 , X_2 and X_3 chosen uniformly from those that satisfy $(X_1, X_2) \in \mathcal{C}_{12} \subseteq A_1 \times A_2$ and $(X_2, X_3) \in \mathcal{C}_{23} \subseteq A_2 \times A_3$.

We claimed that the variables (X_1, X_2, X_3) then satisfy what the book calls the “Markov” property. That is, for all a_1, a_2, a_3 ,

$$P [X_1 X_3 = a_1 a_3 | X_2 = a_2] = P [X_1 = a_1 | X_2 = a_2] P [X_3 = a_3 | X_2 = a_2].$$

I’ll now sketch a proof. It basically follows by recalling the definition of the probability of an event conditioned on $X_2 = a_2$. We first note that the set of choices for (X_1, X_2, X_3) given that $X_2 = a_2$ is

$$S_{a_2} \stackrel{\text{def}}{=} \{(X_1, a_2, X_3) : (X_1, a_2) \in \mathcal{C}_{12} \text{ and } (a_2, X_3) \in \mathcal{C}_{23}\}.$$

Conditioning on $X_2 = a_2$, we obtain a sample chosen uniformly from S_{a_2} . Thus, for $(a_1, a_2, a_3) \in S_{a_2}$,

$$P [(X_1, X_2, X_3) = (a_1, a_2, a_3) | X_2 = a_2] = \frac{1}{|S_{a_2}|}.$$

Note that

$$|S_{a_2}| = |\{a_1 : (a_1, a_2) \in \mathcal{C}_{12}\}| |\{a_3 : (a_2, a_3) \in \mathcal{C}_{23}\}|$$

The claim now follows from observing that

$$P [(X_1, X_2) = (a_1, a_2) | X_2 = a_2] = \frac{|\{a_3 : (a_2, a_3) \in \mathcal{C}_{23}\}|}{|S_{a_2}|} = \frac{1}{|\{a_1 : (a_1, a_2) \in \mathcal{C}_{12}\}|}$$

and

$$P [(X_2, X_3) = (a_2, a_3) | X_2 = a_2] = \frac{|\{a_1 : (a_1, a_2) \in \mathcal{C}_{12}\}|}{|S_{a_2}|} = \frac{1}{|\{a_3 : (a_2, a_3) \in \mathcal{C}_{23}\}|}.$$

7.2 Simplifying Probability Computation

First, let's establish that the fundamental quantities we are interested in have the form

$$P [X_i = a_i | E],$$

where E is some event, usually a union of the observed variables. We will typically want these values for all a_i , so we really want a vector

$$(P [X_i = a_1 | E], P [X_i = a_2 | E], \dots, P [X_i = a_n | E]),$$

where a_1, \dots, a_n are the symbols in the alphabet A_i . We will denote such a vector by

$$\vec{P} [X_i | E].$$

Using this notation, and letting \odot denote componentwise product $((a, b) \odot (c, d) = (ac, bd))$, we have

$$\vec{P}^{post} [X_i | E] = \vec{P}^{prior} [X_i] \odot \vec{P}^{ext} [X_i | E]$$

Before returning to our probability computations for (X_1, X_2, X_3) , I'll also introduce the simpler notation $P^{ext} [X_i = a_i | Y_i]$ for $P^{ext} [X_i = a_i | Y_i = b_i]$. We will use this notation whenever b_i is fixed throughout our computation, which it generally is as it is what was received.

We then have, from Lemma 6.2.1,

$$\vec{P}^{post} [X_1 | Y_1 Y_2 Y_3] = \vec{P}^{prior} [X_1] \odot \vec{P}^{ext} [X_1 | Y_1] \odot \vec{P}^{ext} [X_1 | Y_2 Y_3],$$

and, from Lemma 7.0.1,

$$P^{ext} [X_1 = a_1 | Y_2 Y_3] = \sum_{a_2: (a_1, a_2) \in \mathcal{C}_{12}} P [X_2 = a_2 | X_1 = a_1] P^{ext} [X_2 = a_2 | Y_2] P^{ext} [X_2 = a_2 | Y_3].$$

Using these formulas, we go through the following steps to compute $\vec{P}^{post} [X_1 | Y_1 Y_2 Y_3]$.

1. Compute $\vec{P}^{ext} [X_2 | Y_3]$. This computation only depends upon Y_3 , and comes from the formula:

$$P^{ext} [X_2 = a_2 | Y_3] \sim \sum_{a_3: (a_2, a_3) \in \mathcal{C}_{23}} P [X_3 = a_3 | X_2 = a_2] P [Y_3 | X_3 = a_3].$$

2. Compute $\vec{P}^{ext} [X_2 = | Y_2]$, and compute

$$\vec{P}^{ext} [X_2 = | Y_2] \odot P^{ext} [X_2 = a_2 | Y_3].$$

3. For each a_1 , compute

$$\vec{P} [X_2 | X_1 = a_1] \odot \vec{P}^{ext} [X_2 | Y_2] \odot \vec{P}^{ext} [X_2 | Y_3],$$

and then sum the resulting vector.

4. Take the output of the previous step, and \odot product it with $\vec{P}^{prior} [X_1] \odot \vec{P}^{ext} [X_1 | Y_1]$.

If you look at the flow of this computation, it can be understood as a vector being passed from X_3 to X_2 between steps 1 and 2, and a vector being passed from X_2 to X_1 between steps 3 and 4.