By Tijmen Tieleman.

This page was written as a more technical complement to video 13b, about belief nets. First, watch the video, then read this, and then try to understand how the two are two different ways of telling the same story.

Like a Boltzmann Machine (BM), a Sigmoid Belief Network (SBN) is a generative model, i.e. it describes a probability distribution over data cases (data vectors) in some data space. Recall that for a BM, that distribution involves energies and a normalization constant. For an SBN, the distribution is defined quite differently.

In the video, you heard the informal explanation of what an SBN probability distribution looks like, described as a generative process ("to generate data, first we do this, then this, then ..."). Here's the formal description.

Like a BM, an SBN consists of many binary units, some visible and some hidden. BMs and SBNs are both graphs, but the difference is that an SBN is a directed (and acyclic) graph, whereas a BM is an undirected graph. For SBNs, the convention is that the leaf nodes are the visible ones, and are drawn at the bottom of the image. A directed connection from node A to node B means that node A is a parent of node B. Nodes can have any number of parents, including zero. The top nodes have zero parents, so let's call them orphans (I don't know if that's the standard name for them).

The weight on the connection from parent j to child i is called w_{ji}, and the bias of node i is called b_i.

At the heart of the SBN probability distribution is the probability of a unit turning on, conditional on the state of its parents. For node i with state s_i and set of parents Pa_i, that conditional probability of the node turning on (i.e. $s_i = 1$) is as follows: $P(s_i = 1 | s_{\mathrm{Pa}_i}) = \sigma\left(b_i + \sum j \in \mathrm{Pa}_i s_j \cdot wji\right)$, where σ is the standard logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$. Here, $s\mathrm{Pa}_i$ stands for the state of all the parents of node i. Of course, the conditional probability of the unit turning off is simply one minus the conditional probability of the unit turning on.

If node i is an orphan, then $\mathrm{Pa}_i = \emptyset$. In that case, the formula $P(s_i = 1 | s\mathrm{Pa}_i) = \sigma\left(b_i + \sum j \in \mathrm{Pa}_i s_j \cdot wji\right)$ simplifies to $P(s_i = 1) = \sigma(b_i)$. Notice how that conditional probability turned into an unconditional one, because the conditioning is on the state of the parents, and there are no parents, so really we're not conditioning on anything. In other words, an orphan's probability of turning on depends only on its bias.

Now, here's the probability of a full configuration, i.e. a specific state for every unit, visible and hidden. This is the main component of the formal definition of an SBN. Notice how different it is from the probability of a full configuration for a BM.

$P(s) = \prod_i P(s_i | s\mathrm{Pa}_i)$

The last part of the story is the same as it is for BMs: the probability of a specific state for just the visible units is the sum of probabilities of all full configurations in which the visible units are in that specific state. In other words, we're summing over all configurations of the hidden units.

We can also express that in mathematical notation, and it'll be exactly the same as it was for BMs. We're using $\mathbf{s_v}$ to denote the specific state that we care about of all visible units. I'm writing $\mathbf{s_v}$ in bold face to emphasize that it is a collection of numbers (one could say a vector), as opposed to a single number. We're using $\mathbf{s_h}$ to denote a configuration of all hidden units, and the $\sum \mathbf{s_h}$ is a sum over all $2^{n_{\text{hid}}}$ such configurations. Then $P(\mathbf{s_v}) = \sum \mathbf{s_h} \mathbf{P}(\mathbf{s_v}, \mathbf{s_h})$, where $P(\mathbf{s_v}, \mathbf{s_h})$ is the probability of a completely specified full configuration, which is defined above (under the name $P(s)$).

It follows from the above definitions that the orphans are independent of each other. However, conditional on the states of one or more of their descendants, they're typically no longer independent of each other, as you'll be verifying in quiz 13.

Now see if you can connect this description to the informal description that you heard in the video. If you understand that, you'll have a very good starting point for the next video (about learning SBNs) and the quiz (which involves doing these calculations with some specific numbers).