



Why

1

Definition

Let X and Y be sets. A *hidden memory chain* (*hidden markov chain*, *hidden memory model*, *hidden markov model*², *HMM*) with *hiddens* (or *latents*) Y and *observations* Y of length n is a joint distribution $p : X^n \times Y^n \rightarrow [0, 1]$ satisfying

$$p(x, y) = f(x)g(y_1, x_1) \prod_{i=1}^n h(x_i, x_{i-1})g(y_i, x_i),$$

where $f : X \rightarrow [0, 1]$ is a distribution, and g and h are functions satisfying $g(\cdot, \xi)$ and $h(\cdot, \xi)$ are distributions on Y and X , respectively, for all $\xi \in X$.

Proposition 1. *p so defined is a distribution. The function f is the distribution p_1 . For all $i = 1, \dots, n$, $p_{n+1|i} \equiv gg$. For all $i = 2, \dots, n$, $p_{i|i-1} \equiv h$.*³

Clearly, $p_{1,\dots,n}$ is a memory chain (see **Memory Chains**). For this reason, we continue to refer to h as the *conditional distribution*. We continue to refer to f as the *initial distribution*. We refer to h as the *observation distribution*.

The word “hidden” refers to the situation in which we observe outcomes y , and we hypothesize that they were “generated by” unobserved outcomes x .

¹Future editions will include.

²This term is universal. We avoid it because of the Bourbaki project’s policy on naming. The skeptical reader will note (as in **Memory Chains**) that our term and this term have the same initials.

³Future editions will define everything needed in this proposition in the proposition statement, as opposed to saying “ p so defined”.

