



Why

We consider a simple distribution on sequences.¹

Definition

A *memory chain* (*markov chain*², *memory model*, *markov model*) on a set A of length n is a joint distribution $p : A^n \rightarrow [0, 1]$ satisfying

$$p(a) = f(a_1) \prod_{i=2}^n g(a_i, a_{i-1})$$

for some distribution $f : A \rightarrow [0, 1]$ and $g : A^2 \rightarrow [0, 1]$ is a function for which $g(\cdot, \alpha) : A \rightarrow [0, 1]$ is a distribution on A for every $\alpha \in A$.

Proposition 1. *p so defined is a distribution. The function g is the conditional distribution $p_{i|i-1}$ for $i = 2, \dots, n$. The function f is the first marginal p_1 .*

For this reason, we call g the *conditional distribution* of the Markov chain. We call f the *initial distribution*.

Terminology

Let $A = \prod_i A_i$ and $p : A \rightarrow [0, 1]$ be a distribution. On one hand, p is said to be *memoryless* (or have *zero-order mem-*

¹Future editions will modify and expand.

²This term is universal. We avoid it in these sheets because of the Bourbaki project's policy on naming. The skeptical reader will note, at least, out term and this term terms have the same initials.

ory) if $p = \prod_i p_i$. In particular, $p_{i|i-1}(\alpha, \beta) = p_i(\alpha)$ for every $i = 2, \dots, n$. On the other hand, p is said to have *first-order memory* if $p = p_1 \prod_{i=2}^n p_{i|i-1}$

If, in addition, $A_i = A_1$ for all $i = 1, \dots, n$, then we say that p has *homogenous conditionals* if $p_{i|i-1} = p_{j|j-1}$ for $i \neq j = 2, \dots, n$. In this language, a memory chain is a joint distribution with first-order memory and homogenous conditionals.

