1 Model Framework

We are interested in models with asymptotically power-law decay of the mutual information measure with respect to the distance between the tokens in the sequence. So far so good. But what does it *actually* mean?

The tokens, represented by random variables X_t, are elements of a finite alphabet Σ. The distance between X_t and $X_{t+\tau}$ is τ, and for every t and every τ we want to bound

$$I(X_t, X_{t+\tau}) \in \Omega(\tau^{-\alpha}),\ I(X_t, X_{t+\tau}) \in \mathcal{O}(\tau^{-\beta}) \quad ,$$

for some fixed $\alpha, \beta \in \mathbb{R}_{>0}$. The first condition is the important one, while the latter ensures that $I(X_t, X_{t+\tau}) \xrightarrow{\tau\to\infty} 0$. We also may replace the latter condition by this one.

This was straight forward. The challenging part is to define what a model is. In the case of Markov chains this seems trivial: We define a finite set of parameters (the transition probabilities), and we get a model over Σ^*, that is for every $n \in \mathbb{N}$ the model defines a probability measure over Σ^n. Thus, the first conclusion is every model S must define a probability measure over Σ^n for every $n \in \mathbb{N}$.

As a first formalization, S is a function $S : (n, w) \mapsto [0, 1]$, for $n \in \mathbb{N},\ w \in \Sigma^n$ s.t. $\sum_{w\in\Sigma^n} S(n, w) = 1$.

But really, we want to restrain S in order to have reasonable time and space complexity, and to ensure the model is *reasonable*, which means that the language of $S_n(w)$ should look *similar* to $S_{n+d}(w)$, whatever this might mean, where we used the notation $S_n(w) \equiv S(n, w)$. We also write w_i for X_i. Really, w is a 1-indexed String of X_i.

We present one strict definition for this *similarity* in the following definition:

Definition 1.1. We say S is *well behaved* iff for every $n \in \mathbb{N},\ w \in \Sigma^{n+1}$ it holds true that

$$\sum_{w_{n+1}\in\Sigma} S_{n+1}(w) = S_n(w_{-(n+1)}) \quad .$$

Now, we want to look at how we might restrict our model $(S_n)_{n\in\mathbb{N}} \equiv S$. One approach might be to define a model structure for every $n \in \mathbb{N}$ with parameters $\boldsymbol{\theta}$, thus $S_n \in \{S_n(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$. We write $S_{n,\boldsymbol{\theta}}$ for S_n with parameters $\boldsymbol{\theta}$.