Non stochastic sources

- Empirical 0^{th}-order entropy of string x
  $= H(x) = H_0(x) = \sum_{i=1}^{|\Sigma|} p_i \lg 1/p_i$ where $p_i = f_i/n$.
  — misses correlation

- Empirical k^{th}-order entropy
  $H_k = \sum_{|w|=k} \underbrace{\Pr[w \text{ occurring}]}_{= \frac{\#\text{ occurrences of } w}{n}} \times H_0(\underbrace{\text{successors}(w)}_{\text{string of next letter after each occurrence of } w \text{ (order is irrelevant)}})$

- $|x| H_k(x)$ is a lower bound on compression by k^{th} order code — code depending on last k letters
- $H_{k+1} \le H_k \quad \forall k \ge 0$

Realistic? $H_k(aaaaaaaaaa) = 0$. !

- modified entropy $H_0^*(x) = \begin{cases} 0 & \text{if } |x|=0 \\ (1+\lg|x|)/|x| & \text{if } |x|\neq 0 \text{ but } H_0(x)=0 \\ H_0(x) & \text{else} \end{cases}$
  ↳ minimal fix!

- oops: now depending on more chars. can actually hurt (e.g. long string of a's)
- define H_k^* to allow depending on shorter strings — anything $\le k$ & take overall min.