# Universal Coding in Order Models

Henrik Densing Petersen

May 29, 2008

## Acknowledgements


- Co-authors Flemming Topsøe and Peter Harremöes. The former especially for support and guidance.
- Also, thanks to Lars Lau Hansen for shouldering my teaching responsibilities while I'm away.

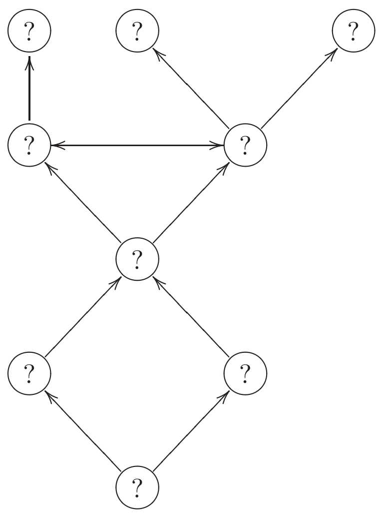




Figure 1: An ordering on a 9 letter alphabet. Can the universal code be determined wihtout brute force trial and error?

## SETUP

- Let $\mathbb{A}$ be a finite set and consider a subset $\mathcal{P}$ of $M_+^1(\mathbb{A})$.
- Generally, we refer to $\mathcal{P}$ as our model, and we wish to determine a probability distribution $P^*$ on $\mathbb{A}$ such that

$$\sup_{P \in \mathcal{P}} D(P \| P^*) = \inf_{Q \in M_+^1(\mathbb{A})} \sup_{P \in \mathcal{P}} D(P \| Q)$$

  where $D$ is the Kullback-Leibler divergence. When such a distribution exists, we say that $P^*$ is the minimax predictor for the model $\mathcal{P}$.

| The goal is to determine, in closed form, such a minimax predictor for as many explicit (families of) models as possible. |
|---|

## ORDER MODELS

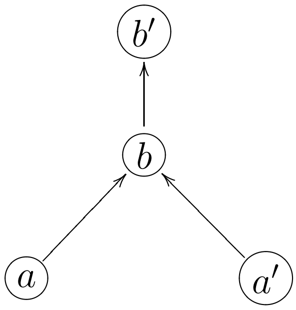
- Of specific interest are order consistent models,. Consider a (usually partial) ordering '$\leq$' on $\mathbb{A}$ and let $\mathcal{P}$ be the model of probability distributions on $\mathbb{A}$ satisfying

$$a \leq b \Rightarrow P(a) \geq P(b)$$

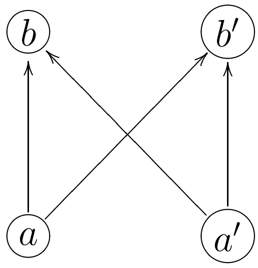
- Pictorially we depict an ordering on $\mathbb{A}$ by a graph where ⓐ⟶ⓑ means that $a \leq b$, and we leave out arrows that are redundant due to transitivity.
- It can be proved that for order models, the minimax predictor always exists, and it is unique.

## SIMPLE EXAMPLES

- Consider the alphabet $\mathbb{A} = \{a, a', b, b'\}$. On this we may consider, among others, 2 order models given by the graphs below.




$\mathcal{P}_1$




$\mathcal{P}_2$

- Remarkably, it turns out that the minimax predictors are equal for the two models. Specifically, in both cases, $P^*(a) = P^*(a') = \frac{16}{34}$ and $P^*(b) = P^*(b') = \frac{1}{34}$.
  This is not so hard to see, once we have noticed that for $\mathcal{P}_1$, the minimax predictor assigns equal probability to $b$ and $b'$. (to see this, of course we need to actually calculate it, but this is doable!)

## BASIC TECHNIQUES

- Our basic strategy for determining minimax predictors for order models is to make (sometimes educated) guesses and then apply the following
- Kuhn-Tucker criterion: If $P^*$ is a (finite) convex mixture $P^* = \sum \alpha_k P_k$ of order consistent distributions, and

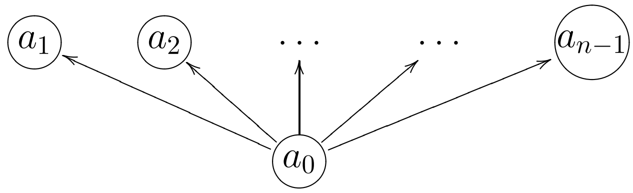
$$D(P_k \| P^*) = R \quad \text{for all } k,$$
$$D(P \| P^*) \leq R \quad \text{for all } P \in \mathcal{P},$$

then $P^*$ is the minimax predictor for $\mathcal{P}$.

- The uniqueness result mentioned above is in fact even more powerful, as it says that there always exists some set of order consistent distributions/weights yielding a minimax predictor through the Kuhn-Tucker criterion.
- A set $\Delta \subseteq \mathbb{A}$ is said to be hereditary if whenever $a \in \Delta$, we have $a^{\downarrow} \subseteq \Delta$, where $a^{\downarrow} = \{b \in \mathbb{A} \mid b \leq a\}$.
- The search for a set of distributions and weights for application of the Kuhn-Tucker criterion is decilitated by the fact that any order consistent distribution is a convex combination of uniform distributions supported on connected hereditary subsets of $\mathbb{A}$. Using convexity of the Kullback-Leibler divergence it is then realized that we need only look at convex combinations of such uniform distributions.

## OVERVIEW OF CURRENT RESULTS

**Trees of height 1**. For a tree of height 1 as below,




the minimax predictor is given by $P^*(a_1) = \cdots = P^*(a_{n-1})$ for all $n$, and

$$P^*(a_0) = \begin{cases} \frac{1}{k} & \text{when } a_k \le n \le b_k \\ 1 - \frac{n-1}{n-1+(1+\frac{1}{k-1})^{(k-1)k}} & \text{when } b_{k-1} \le n \le a_k \end{cases}$$

for certain sequences $(a_k)$ and $(b_k)$ satisfying $1 = b_1 < a_2 < b_2 < a_3 < b_3 < \cdots$.

**Matrix orderings**. Suppose that $\mathbb{A} = \{a_{ij} \mid i, j = 1, \ldots n\}$. The ordering given by $a_{ii} \le a_{ij}, a_{ji}$ for all $i, j$ and no other relations is a (2-dimensional) matrix ordering. Again there is an exact result determining the minimax predictor, and again there is a peculiar variation with $n$.

**Co-trees**. Suppose that the ordering on $\mathbb{A}$ defines a (suspended) co-tree, i.e. that there is a unique maximal element and that each element has at most 1 immediate successor. It is then possible to determine the minimax predictor by an algorithm which terminates in polynomial time (in the cardinality of $\mathbb{A}$).

## COMMENTS ON OPEN PROBLEMS

**More specific results for order structures**. Given the peculiar variation with $n$ of the minimax predictors for tree / matrix models, it seems unlikely that there is a general algorithm to cover all order structures. At the very least, it indicates that this would be rather complicated. However, There are certainly more specific order models for which the monimax predictor can be determined, either in closed form or by a terminating algorithm.

**Relation of order models to other models**.The matrix and tree orderings are interesting not only in their own right, but also because they are connected to Bernoulli sources. It turns out that the matrix ordering above have the same minimax predictor as the problem of predicting the first 2 characters from an $n$-letter Bernoulli source with no previous knowlegde, and that trees of height 1 describe the problem of predicting 1 character, already knowing what came immediately before it. Can this be extended? Are there other similar cases?

**General structural results**. The remark concerning the calculation of the minimax predictor for $\mathcal{P}_2$ on the 'SIMPLE EXAMPLES' sheet purely from the knowlegde of the minimax predictor for $\mathcal{P}_1$ can be extended and generalized. Lacking an algorithm for general order structures, such general tricks could prove very effective, or at least an amusing curiosity.