Deterministic Compression with Uncertain Priors

Presenting: Elad Haramaty. Joint with: Madhu Sudan

Main Question

Knowledge
- Sender input:
  - probability \( P \)
  - message \( m \) in \( \{1, \ldots, N\} \), drawn according to \( P \)
- Receiver input:
  - probability \( Q \) “close” to \( P \)

Game Rules
- Sender sends \( k \) bits to the receiver.
- Receiver tries to guess \( m \).

Goal
- Minimize the amount of communication such that the receiver is able to know \( m \).

Easier problem
- Bob has a very good estimation of Alice’s top boy ranking (for each boy he can be wrong by at most one position).
- Alice tries to communicate to Bob her top ranked boy.

Motivation
- Communication between two parties who estimate the distribution of the messages independently.
- Explain phenomena in human communication:
  - Dictionary: Words often have multiple meanings.
  - Redundant: But not as in any predefined way (not an error-correcting code).

schemes

**No assumption on \( P \) & \( Q \)**
- Best possible performance: \( \log(N) \)
- Scheme: binary representation

**\( P = Q \)**
- Best possible performance: \( H(P) = \sum_{m=1}^{N} P \log \left( \frac{1}{P} \right) \)
- Scheme: Huffman coding

**\( P \sim Q \), high entropy**
- Sender:
  1. Send an estimation of \( P_m \).
  2. Find a function \( h \) such that \( h(m) \) is different from \( h(m') \) for any \( m' \) that has a close distribution to \( m \).
  3. Send \( h \)'s identifier.
  4. Send \( h(m) \).
- Decoder:
  1. Output message \( m^* \) such that \( Q_{m^*} \) is close to \( P_m \) and \( h(m^*) = h(m) \)
- Performance: \( O(H(P) + \log \log(N)) \)
- Correctness:
  - Need to prove: There is a “small” family \( \{h\} \) of functions to a “small range” that for any \( m \) and a set of \( \{m'\} \), there is a function in the family which satisfies step 2’s requirements.

**\( P \sim Q \), low entropy**
- Scheme: “Chain Coloring”
- Performance: \( \exp(H(P) + \log(N)) \)