Compression under uncertain priors

Madhu Sudan
Microsoft, New England

Based on joint works with:
Adam Kalai (MSR), Sanjeev Khanna (U.Penn) and Brendan Juba (Harvard)

Motivation: Human Communication

- Human communication (dictated by languages, grammars) very different.
  - Grammar: Rules, often violated.
  - Dictionary: Often multiple meanings to a word.
  - Redundant: But not as in any predefined way (not an error-correcting code).

- Theory?
  - Information theory?
  - Linguistics? (Universal grammars etc.)?

Behavioral aspects of natural communication

- (Vast) Implicit context.
- Sender sends increasingly long messages to receiver till receiver "gets" (the meaning of) the message.
- Sender may use feedback from receiver if available; or estimates receiver's knowledge if not.
- Language provides sequence of (increasingly) long ways to represent a message.
- Question: What is the benefit of choosing short/long messages?

Model:

- Reason to choose short messages: Compression.
  - Channel is still a scarce resource; still want to use optimally.

- Reason to choose long messages (when short ones are available): Reducing ambiguity.
  - Sender unsure of receiver's prior (context).
  - Sender wishes to ensure receiver gets the message, no matter what its prior (within reason).

Model

- Wish to design encoding/decoding schemes (E/D) to be used as follows:
  - Sender has distribution P on M = {1,2,...,N}
  - Receiver has distribution Q on M = {1,2,...,N}
  - Sender gets $X \in M$
  - Sends E(P,X) to receiver.
  - Receiver receives Y = E(P,X)
  - Decodes to $\hat{X}$ = D(Q,Y)

  - Want: X = $\hat{X}$ (provided P,Q close),
    - While minimizing $Exp_{X \leftarrow P}\,|E(P,X)|$

Contrast with some previous models

- Universal compression?
  - Doesn't apply: P,Q are not finitely specified.
  - Don't have a sequence of samples from P; just one!
- K-L divergence?
  - Measures inefficiency of compressing for Q if real distribution is P.
  - But assumes encoding/decoding according to same distribution Q.
- Semantic Communication:
  - Uncertainty of sender/receiver; but no special goal.

Closeness of distributions:

- P is α-close to Q if for all $X \in M$,

$$\frac{1}{\alpha} \leq \frac{P(X)}{Q(X)} \leq \alpha$$

- P α-close to Q $\quad \Rightarrow \quad D(P||Q), D(Q||P) \leq \log \alpha$.

Dictionary = Shared Randomness?

- Modelling the dictionary: What should it be?
- Simplifying assumption – it is shared randomness, so …
- Assume sender and receiver have some shared randomness R and X is independent of R.
  - Y = E(P,X,R)
  - $\hat{X}$ = D(Q,Y,R)
- Want $\forall X,\ \Pr_R\left[\hat{X} = X\right] \geq 1 - \epsilon$

Solution (variant of Arith. Coding)

- Use R to define sequences
  - $R_1\ [1], R_1\ [2], R_1\ [3], \ldots$
  - $R_2\ [1], R_2\ [2], R_2\ [3], \ldots$
  - …
  - $R_N\ [1], R_N\ [2], R_N\ [3], \ldots.$
- $E_\alpha(P, x, R) = R_x[1 \ldots L]$, where L chosen s.t. $\forall z \neq x$

  Either $R_z[1 \ldots L] \neq R_x[1 \ldots L]$

  Or $P(z) < \frac{P(x)}{\alpha^2}$
- $D_\alpha(Q, y, R) = \hat{x}\ s.t.\ \hat{x}$ max $Q(\hat{x})$ among $\hat{x} \in \{z | R_z[1 \ldots L] = y\}$

Performance

- Obviously decoding always correct.
- Easy exercise:
  - $\mathrm{Exp}_X \; [E(P, X)] = H(P) + 2 \log \alpha$
- Limits:
  - No scheme can achieve $(1 - \epsilon) \cdot [H(P) + \log \alpha]$
  - Can reduce randomness needed.

Implications

- Reflects the tension between ambiguity resolution and compression.
  - Larger the α ((estimated) gap in context), larger the encoding length.
- Coding scheme reflects the nature of human process (extend messages till they feel unambiguous).
- The "shared randomness" is a convenient starting point for discussion
  - Dictionaries do have more structure.
  - But have plenty of entropy too.
  - Still … should try to do without it.

Future work?

- Upcoming:
  - Some partial derandomization
    - [w. E. Haramaty and G.Ranade]
    - Neat connections to fractional graph chromaticity and the Knesser conjecture/Lovasz theorem.
- Needed:
  - Better understanding of forces on language.
    - Information-theoretic
    - Computational
    - Evolutionary

Thank You