Likelihood Encoder for Source Coding

Paul Cuff, Eva C. Song

Likelihood Encoder

- Given:
  - Joint Distribution: $P_{X,U}$
  - Codebook: $\{u^n(m)\}_{m=1}^{2^{nR}}$

- Encoder:

$$P_{M|X^n}(m|x^n) \propto \prod_{t=1}^{n} P_{X|U}(x_t|u_t(m))$$

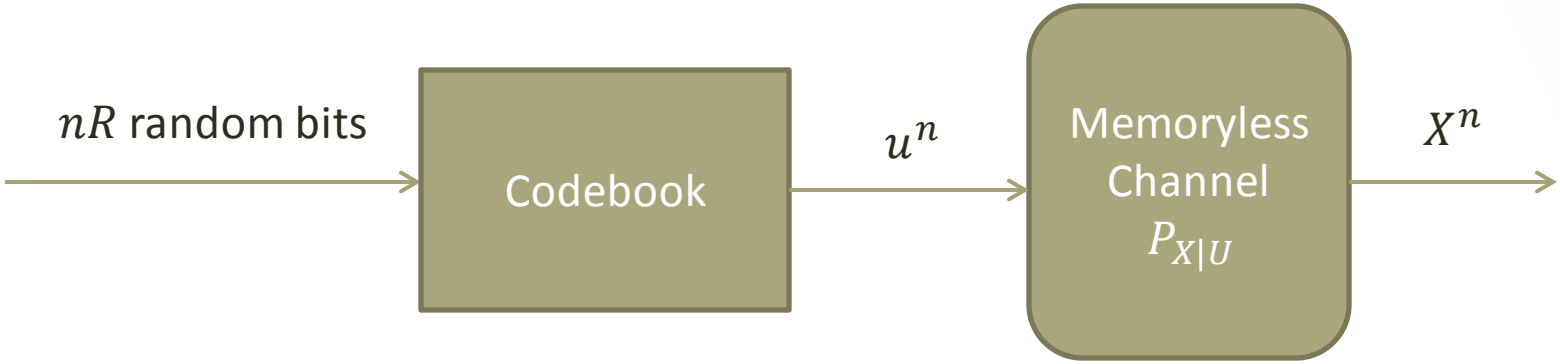
- Encoder (given a codebook and joint distribution) chooses messages stochastically, with probability proportional to the likelihood of the observation X^n given each codeword.

Previous uses of likelihood encoder

- Our Research Group
  - Schieler, Satpathy, Song
  - Purposes:
    - Channel synthesis
    - Rate distortion theory for secrecy systems

- Likelihood Decoder proposed for channel decoding
  - [Yassaee, Aref, Gohari 13], ISIT Student Paper Award

Soft Covering Lemma



- Lemma: If the codebook is randomly constructed and $R > I(X;U)$, then

$$\boldsymbol{E}_C \left\| P_{X^n} - \prod P_X \right\|_{TV} \to 0$$

- See
  - [Wyner 75]: Common Information
  - [Han-Verdu 93]: Channel Resolvability
  - [Cuff 13]: Distributed Channel Synthesis

Comment

- Analysis of likelihood encoder is simple. If rates are high enough, induced distribution behaves as a uniform distribution over the codebook connected to the channel.

Rate Distortion Theory [Shannon]

i.i.d. source sequence X^n distributed according to $X_i \sim \overline{P}_X$

Encoder $f_n : \mathcal{X}^n \mapsto \mathcal{M}$ (possibly stochastic).
Decoder $g_n : \mathcal{M} \mapsto \mathcal{Y}^n$ (possibly stochastic).
Compression rate: R, i.e. $|\mathcal{M}| = 2^{nR}$.
Fidelity requirement: $\mathbb{E} \quad d(X^n, Y^n) \quad \leq \quad D$, where $d(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, y_i)$.

$$R = \min_{P_{Y|X}:\mathbb{E}[d(X,Y)]\leq D} I(X;Y).$$

Proof of Achievability

- Select $\bar{P}_{X,Y} = \bar{P}_X \bar{P}_{Y|X}$ such that
  - $R > I(X;Y)$
  - $\boldsymbol{E}[d(X,Y)] < D$

- Use a random codebook and the likelihood encoder

Proof Continued

- Induced distribution: $P^{(0)}_{X^n,Y^n}$
- Idealized distribution: $P^{(1)}_{X^n,Y^n}(x^n, y^n)$

$$\triangleq \left(\prod_{t=1}^{n} \overline{P}_{X|Y}(x_t|y_t)\right) \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \mathbb{1}\{Y^n(m) = y^n\}.$$

- Notice:
  - $\boldsymbol{E}_C P^{(1)}_{X^n,Y^n}(x^n, y^n) = \prod_{t=1}^{n} \bar{P}_{X,Y}(x_t, y_t)$
  - $P^{(0)}_{Y^n|X^n} = P^{(1)}_{Y^n|X^n}$

Proof Continued

- By the soft covering lemma
  - $\boldsymbol{E}_C \left\| P_{X^n}^{(0)} - P_{X^n}^{(1)} \right\|_{TV} \to 0$
- Therefore
  - $\boldsymbol{E}_C \left\| P_{X^n,Y^n}^{(0)} - P_{X^n,Y^n}^{(1)} \right\|_{TV} \to 0$

$$\begin{aligned}
\boldsymbol{E}_C \boldsymbol{E}_{P^{(0)}} d(X^n, Y^n) &\leq \boldsymbol{E}_C \left[\boldsymbol{E}_{P^{(1)}} d(X^n, Y^n) + d_{max} \left\| P^{(0)} - P^{(1)} \right\|_{TV} \right] \\
&= \boldsymbol{E}_{\bar{P}} d(X^n, Y^n) + d_{max} \boldsymbol{E}_C \left\| P^{(0)} - P^{(1)} \right\|_{TV} \\
&< D \qquad \text{for } n \text{ large enough.}
\end{aligned}$$

Properties of Total Variation

- $\|P_X - Q_X\|_{TV} \leq \left\|P_{X,Y} - Q_{X,Y}\right\|_{TV}$
- $\|P_X - Q_X\|_{TV} = \left\|P_X P_{Y|X} - Q_X P_{Y|X}\right\|_{TV}$
- $\left|\boldsymbol{E}_P f(X) - \boldsymbol{E}_Q f(X)\right| \leq 2d_{max}\|P - Q\|_{TV}$

- If $\boldsymbol{P}(X \neq \widehat{\boldsymbol{X}}) \leq \varepsilon$, then $\left\|P_{X,Y} - P_{\hat{X},Y}\right\|_{TV} \leq \varepsilon$.

Source Coding with Decoder Side Information

[Wyner-Ziv 76]

- Sources X^n and Y^n i.i.d. according to $\bar{P}_{X,Y}$
- Encoder observes X^n
- Decoder observes Y^n
- Message from encoder to decoder at rate R
- Decoder produces $\hat{X}^n$
- Distortion constraint: $\boldsymbol{E}\, d\left(X^n, \hat{X}^n\right) < D$

$$R > I(X;U|Y)$$

where $U - X - Y$ form a Markov chain
and there exists an f such that $\boldsymbol{E}\, d(X, f(U,Y)) < D$

Proof of Achievability

- Select $\bar{P}_{U,X,Y} = \bar{P}_{X,Y}\bar{P}_{U|X}$ and f such that
  - $R > I(X;U|Y)$
  - $\boldsymbol{E}[d(X, f(U,Y))] < D$

- Use a random codebook with two indices:
  - M_a at rate $R_a > I(X;U|Y)$
  - M_b at rate $R_b < I(Y;U)$
  - $R_a + R_b > I(X,Y;U)$

- Likelihood encoder
- Decoder:
  - Channel Decoder to decode M_b
  - Apply function f

Proof Continued

- $P^{(0)} = \bar{P}_{X^n,Y^n} P_{LE}(m_a, m_b | x^n) P_D(m'_b | m_a, y^n) P_f(\hat{x}^n | m_a, m'_b, y^n)$

- $P^{(1)} = Q_{X^n,Y^n,M_a,M_b} P_D(m'_b | m_a, y^n) P_f(\hat{x}^n | m_a, m'_b, y^n)$
  - Q is the idealized distribution
  - Soft covering lemma says $P^{(0)} \approx P^{(1)}$
  - Under $P^{(1)}$, $\boldsymbol{P}(M_b \neq M'_b) \to 0$ by channel decoding proof.

- $P^{(2)} = Q_{X^n,Y^n,M_a,M_b} P_D(m'_b | m_a, y^n) P_f(\hat{x}^n | m_a, m_b, y^n)$

$$\boldsymbol{E}_C P^{(2)} = \prod \bar{P}_{X,\hat{X}}$$

Proof Intuition

$X^n, \boldsymbol{Y^n}$

$\downarrow$ Likelihood Encoder

$u^n(\boldsymbol{m_a}, m_b)$

Proof Intuition

$X^n, \boldsymbol{Y^n}$

Channel

$u^n(\boldsymbol{m_a}, m_b)$

Total Variation Error:

ε_{SC}

Soft Covering Lemma

Proof Intuition

$X^n, \boldsymbol{Y^n}$

Channel

$u^n(\boldsymbol{m_a}, \boldsymbol{m_b})$

Total Variation Error:

ε_{SC}

ε_{Decode}

Channel Decoding

Proof Intuition

$X^n, \boldsymbol{Y^n}$

Channel

$\boldsymbol{U^n}$

Total Variation Error:

ε_{SC}

ε_{Decode}

Expectation over codebooks

Summary

- Simple Proofs of classical (and new) results using the likelihood encoder.

- Proof techniques easily extend to other problems:
  - E.g. Multi-terminal source coding [Berger-Tung]