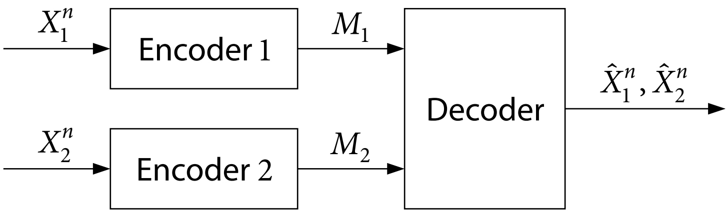
Lecture #3 Distributed Lossless Compression

(Reading: NIT 10.1–10.5, 4.4)

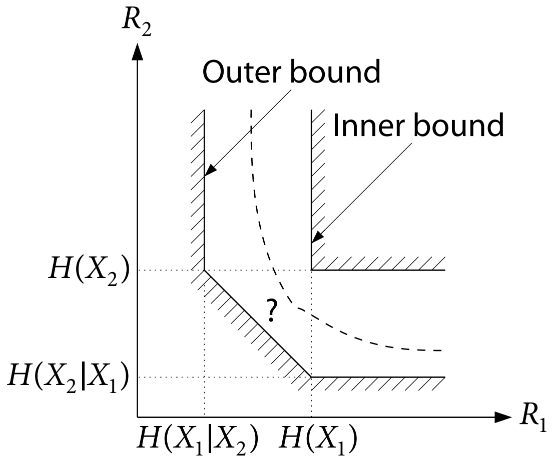
- Distributed lossless source coding
- Lossless source coding via random binning
- Time sharing
- Achievability proof of the Slepian–Wolf theorem
- Extension to more than two sources

Distributed lossless compression system



- Two-component DMS (2-DMS) $(\mathcal{X}_1 \times \mathcal{X}_2, p(x_1, x_2))$
- A $(2^{nR_1}, 2^{nR_2}, n)$ code:
  - Two encoders: $m_1(x_1^n) \in [1:2^{nR_1})$ and $m_2(x_2^n) \in [1:2^{nR_2})$
  - Decoder $(\hat{x}_1^n, \hat{x}_2^n)(m_1, m_2)$
- Probability of error: $P_e^{(n)} = \mathrm{P}\{(\hat{X}_1^n, \hat{X}_2^n) \neq (X_1^n, X_2^n)\}$
- (R_1, R_2) achievable if $\exists$ $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$
- Optimal rate region $\mathscr{R}^*$: Closure of the set of achievable (R_1, R_2)

Bounds on the optimal rate region



- Sufficient condition for individual compression:

$$R_1 > H(X_1), \quad R_2 > H(X_2)$$

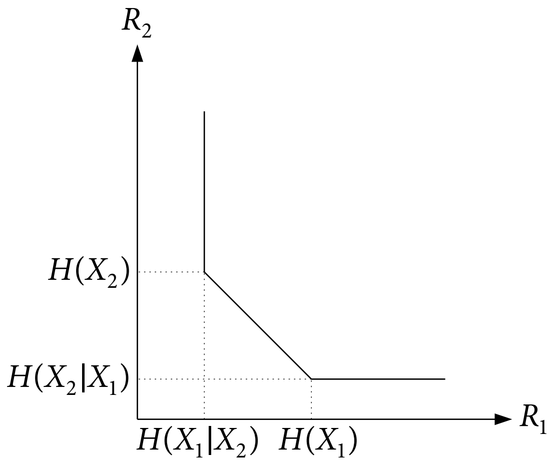
- Necessary condition for centralized compression:

$$R_1 + R_2 \geq H(X_1, X_2)$$

- Can also show that

$$R_1 \geq H(X_1|X_2), \quad R_2 \geq H(X_2|X_1) \quad \text{are necessary}$$

Slepian–Wolf theorem



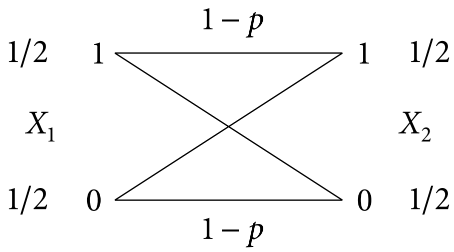
Theorem 10.1 (Slepian–Wolf 1973)

The optimal rate region $\mathscr{R}^*$ is the set of (R_1, R_2) such that

$$R_1 \geq H(X_1|X_2),$$
$$R_2 \geq H(X_2|X_1),$$
$$R_1 + R_2 \geq H(X_1, X_2)$$

Example

- Doubly symmetric binary source (DSBS(p)) (X_1, X_2)

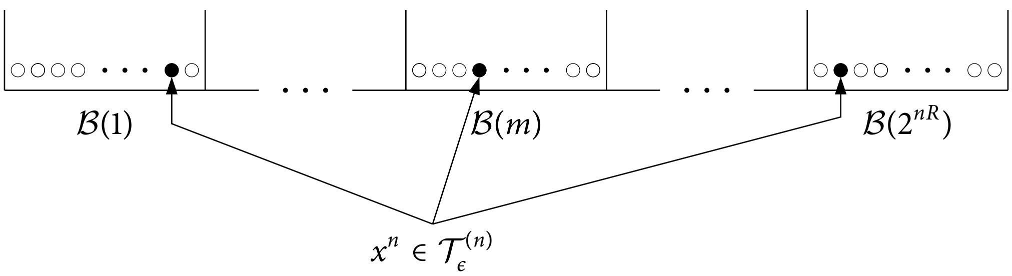


$X_1 \backslash X_2$	0	1
0	$\frac{1-p}{2}$	$\frac{p}{2}$
1	$\frac{p}{2}$	$\frac{1-p}{2}$

- Let $p = 0.01$
- Individual compression: 2 bits/symbol-pair
- Slepian–Wolf coding: $H(X_1, X_2) = 1.0808$ bits/symbol-pair

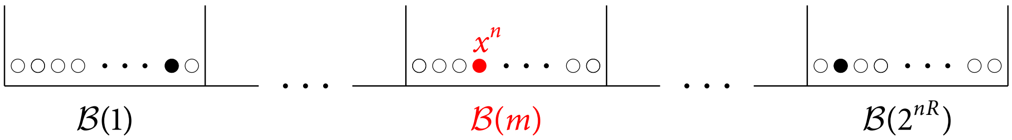
Lossless source coding via random binning

- Codebook generation:
  - Randomly assign an index $m(x^n) \in [1 : 2^{nR}]$ to each sequence $x^n \in \mathcal{X}^n$
  - The set of sequences with the same index m form a bin $\mathcal{B}(m)$, $m \in [1 : 2^{nR}]$
  - Bin assignments are revealed to the encoder and decoder
- Encoding:
  - Upon observing $x^n \in \mathcal{B}(m)$, send the bin index m
- Decoding:
  - Find the unique typical sequence $\hat{x}^n \in \mathcal{B}(m)$



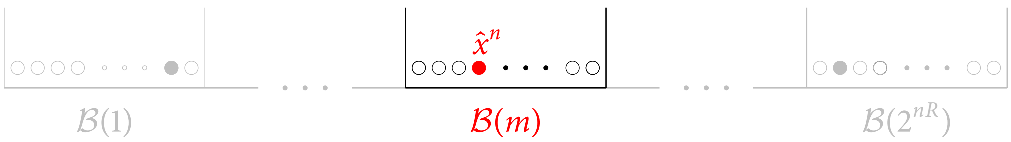
Lossless source coding via random binning

- Codebook generation:
  - Randomly assign an index $m(x^n) \in [1:2^{nR}]$ to each sequence $x^n \in \mathcal{X}^n$
  - The set of sequences with the same index m form a bin $\mathcal{B}(m)$, $m \in [1:2^{nR}]$
  - Bin assignments are revealed to the encoder and decoder
- Encoding:
  - Upon observing $x^n \in \mathcal{B}(m)$, send the bin index m
- Decoding:
  - Find the unique typical sequence $\hat{x}^n \in \mathcal{B}(m)$



Lossless source coding via random binning

- Codebook generation:
  - Randomly assign an index $m(x^n) \in [1:2^{nR}]$ to each sequence $x^n \in \mathcal{X}^n$
  - The set of sequences with the same index m form a bin $\mathcal{B}(m)$, $m \in [1:2^{nR}]$
  - Bin assignments are revealed to the encoder and decoder
- Encoding:
  - Upon observing $x^n \in \mathcal{B}(m)$, send the bin index m
- Decoding:
  - Find the unique typical sequence $\hat{x}^n \in \mathcal{B}(m)$



Analysis of the probability of error

- We bound $P_e^{(n)}$ averaged over random bin assignments
- Let M denote the random bin index of X^n, i.e., $X^n \in \mathcal{B}(M)$
- Note that $M \sim \mathrm{Unif}[1:2^{nR}]$, independent of X^n
- Error events:

$$\mathcal{E}_1 = \{X^n \notin \mathcal{T}_\epsilon^{(n)}\}, \text{ or}$$
$$\mathcal{E}_2 = \{\tilde{x}^n \in \mathcal{B}(M) \text{ for some } \tilde{x}^n \neq X^n,\ \tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}\}$$

Thus

$$\begin{aligned}\mathrm{P}(\mathcal{E}) &\le \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_2)\\ &= \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_2 | X^n \in \mathcal{B}(1))\end{aligned}$$

- By the LLN, $\mathrm{P}(\mathcal{E}_1) \to 0$

Analysis of the probability of error

- Consider

$$\begin{aligned}\mathrm{P}(\mathcal{E}_2 | X^n \in \mathcal{B}(1)) &= \sum_{x^n} \mathrm{P}\{X^n = x^n \,|\, X^n \in \mathcal{B}(1)\}\\ &\quad\cdot \mathrm{P}\{\tilde{x}^n \in \mathcal{B}(1) \text{ for some } \tilde{x}^n \neq x^n, \tilde{x}^n \in \mathcal{T}_\epsilon^{(n)} \,|\, x^n \in \mathcal{B}(1), X^n = x^n\}\\ &\le \sum_{x^n} p(x^n) \sum_{\substack{\tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}\\ \tilde{x}^n \neq x^n}} \mathrm{P}\{\tilde{x}^n \in \mathcal{B}(1) \,|\, x^n \in \mathcal{B}(1), X^n = x^n\}\\ &= \sum_{x^n} p(x^n) \sum_{\substack{\tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}\\ \tilde{x}^n \neq x^n}} \mathrm{P}\{\tilde{x}^n \in \mathcal{B}(1)\}\\ &\le |\mathcal{T}_\epsilon^{(n)}| \cdot 2^{-nR}\\ &\le 2^{n(H(X)+\delta(\epsilon))} 2^{-nR}\end{aligned}$$

- Hence $\mathrm{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R > H(X) + \delta(\epsilon)$

Achievability via linear binning

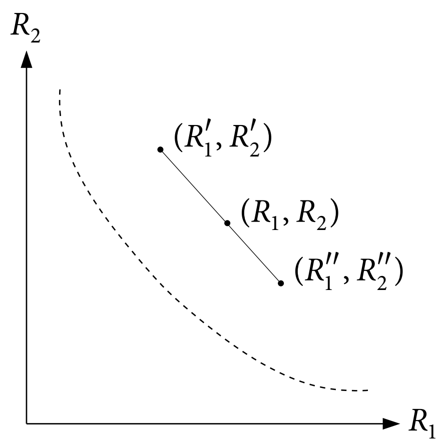
- Let X be a Bern(p) source
- $R = H(X)$ achieved via linear binning (hashing)
  - Let H be a randomly generated $nR \times n$ binary parity-check matrix
  - Encoder sends HX^n
  - Decoder recovers X^n with high probability if $R > H(p)$ (why?)

Time sharing

Proposition 4.1

If $(R_1', R_2'), (R_1'', R_2'') \in \mathscr{R}^*$, then $(R_1, R_2) = (\alpha R_1' + \bar{\alpha} R_1'', \alpha R_2' + \bar{\alpha} R_2'') \in \mathscr{R}^*$ for $\alpha \in [0, 1]$

- The rate region $\mathscr{R}^*$ is convex

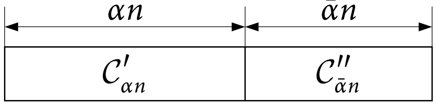


Time sharing

Proposition 4.1

If $(R_1', R_2'), (R_1'', R_2'') \in \mathscr{R}^*$, then $(R_1, R_2) = (\alpha R_1' + \bar{\alpha} R_1'', \alpha R_2' + \bar{\alpha} R_2'') \in \mathscr{R}^*$ for $\alpha \in [0, 1]$

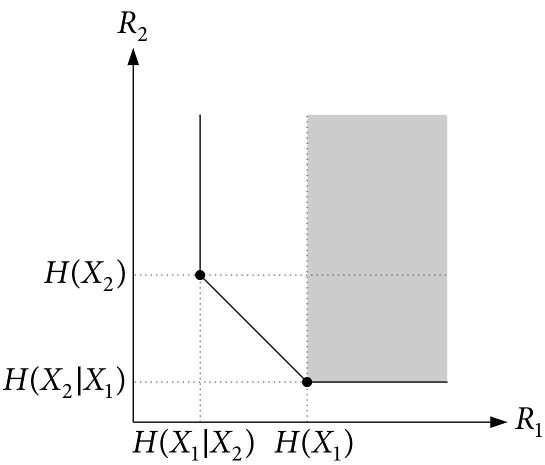
- The rate region $\mathscr{R}^*$ is convex
- Proof: Time sharing argument
  - Let $\mathcal{C}_k'$ be a sequence of $(2^{kR_1'}, 2^{kR_2'}, k)$ codes with $P_{e1}^{(k)} \to 0$
  - Let $\mathcal{C}_k''$ be a sequence of $(2^{kR_1''}, 2^{kR_2''}, k)$ codes with $P_{e2}^{(k)} \to 0$
  - Construct a new sequence of codes by using $\mathcal{C}_{\alpha n}'$ for $i \in [1 : \alpha n]$ and $\mathcal{C}_{\bar{\alpha} n}''$ for $i \in [\alpha n + 1 : n]$



  - By the union of events bound, $P_e^{(n)} \le P_{e1}^{(\alpha n)} + P_{e2}^{(\bar{\alpha} n)} \to 0$
- Remarks:
  - Time division is a special case of time sharing (between $(R_1, 0)$ and $(0, R_2)$)
  - The rate (capacity) region of any source (channel) coding problem is convex

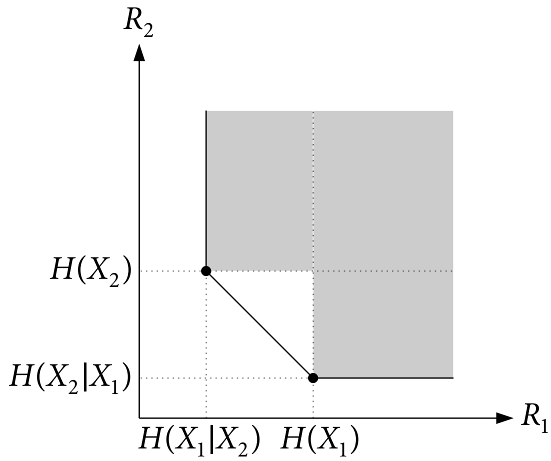
Achievability proof of the S–W theorem (Cover 1975)

- We show that the corner point $(H(X_1), H(X_2|X_1))$ is achievable
- Achievability of the other corner point $(H(X_1|X_2)), H(X_2))$ follows similarly
- The rest of the region is achieved using time sharing



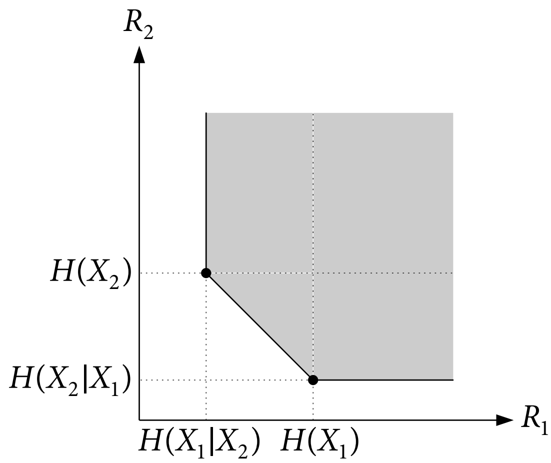
Achievability proof of the S–W theorem (Cover 1975)

- We show that the corner point $(H(X_1), H(X_2|X_1))$ is achievable
- Achievability of the other corner point $(H(X_1|X_2)), H(X_2))$ follows similarly
- The rest of the region is achieved using time sharing



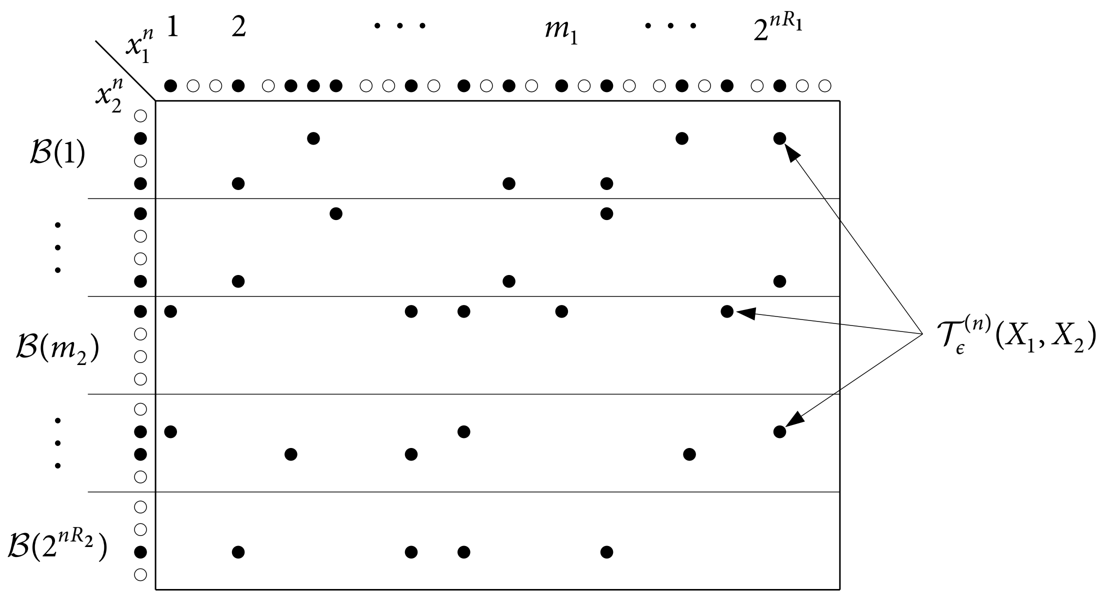
Achievability proof of the S–W theorem (Cover 1975)

- We show that the corner point $(H(X_1), H(X_2|X_1))$ is achievable
- Achievability of the other corner point $(H(X_1|X_2)), H(X_2))$ follows similarly
- The rest of the region is achieved using time sharing



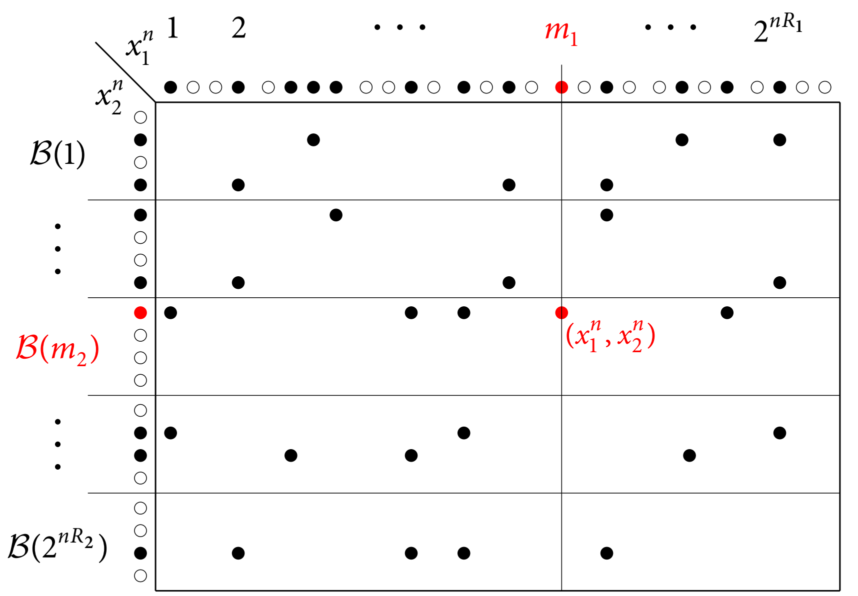
Achievability of $(H(X_1), H(X_2|X_1))$

- Codebook generation:
  - Assign a distinct index $m_1 \in [1:2^{nR_1}]$ to each $x_1^n \in \mathcal{T}_\epsilon^{(n)}(X_1)$, and $m_1 = 1$, otherwise
  - Randomly assign an index $m_2(x_2^n) \in [1:2^{nR_2}]$ to each $x_2^n \in \mathcal{X}_2^n$
  - The sequences with the same index m_2 form a bin $\mathcal{B}(m_2)$



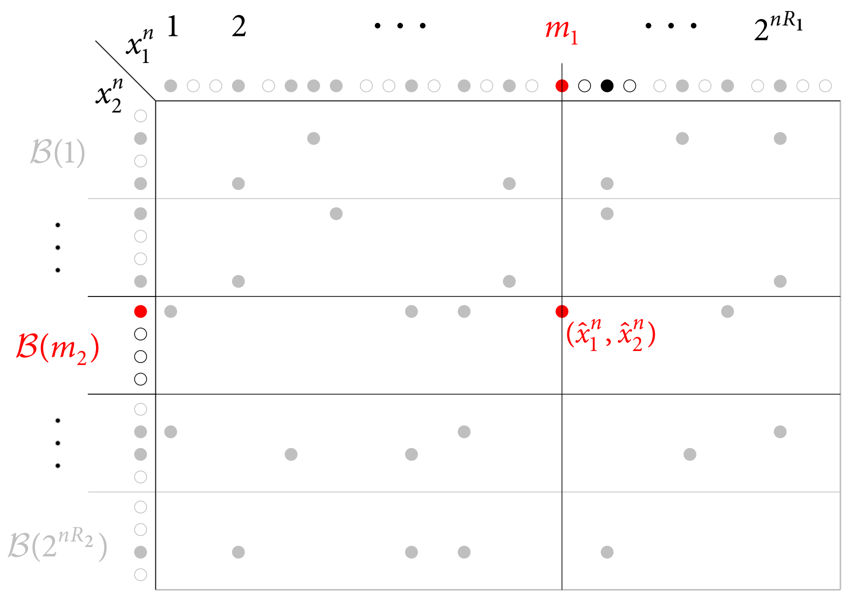
Achievability of $(H(X_1), H(X_2|X_1))$

- Encoding:
  - Upon observing x_1^n, encoder 1 sends the index $m_1(x_1^n)$
  - Upon observing $x_2^n \in \mathcal{B}(m_2)$, encoder 2 sends m_2



Achievability of $(H(X_1), H(X_2|X_1))$

- Decoding: Sources recovered successively
  - Declare $\hat{x}_1^n = x_1^n(m_1)$ for the unique $x_1^n(m_1) \in \mathcal{T}_\epsilon^{(n)}(X_1)$
  - Find the unique $\hat{x}_2^n \in \mathcal{B}(m_2) \cap \mathcal{T}_\epsilon^{(n)}(X_2|\hat{x}_1^n)$
  - If there is none or more than one, the decoder declares an error



Analysis of the probability of error

- Let M_1 and M_2 denote the random bin indices for X_1^n and X_2^n
- Error events:

$$\mathcal{E}_1 = \{(X_1^n, X_2^n) \notin \mathcal{T}_\epsilon^{(n)}\},$$
$$\mathcal{E}_2 = \{\tilde{x}_2^n \in \mathcal{B}(M_2) \text{ for some } \tilde{x}_2^n \neq X_2^n, (X_1^n, \tilde{x}_2^n) \in \mathcal{T}_\epsilon^{(n)}\}$$

  Then,

$$\mathrm{P}(\mathcal{E}) \leq \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_2)$$

- $\mathrm{P}(\mathcal{E}_1) \to 0$ by LLN

Analysis of the probability of error

- By symmetry,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_2) &= \mathsf{P}(\mathcal{E}_2 | X_2^n \in \mathcal{B}(1)) \\
&= \sum_{(x_1^n, x_2^n)} \mathsf{P}\{(X_1^n, X_2^n) = (x_1^n, x_2^n) \,|\, X_2^n \in \mathcal{B}(1)\} \\
&\qquad \cdot \mathsf{P}\{\tilde{x}_2^n \in \mathcal{B}(1) \text{ for some } \tilde{x}_2^n \neq x_2^n, (x_1^n, \tilde{x}_2^n) \in \mathcal{T}_\epsilon^{(n)} \,\big|\, x_2^n \in \mathcal{B}(1), \\
&\qquad\qquad (X_1^n, X_2^n) = (x_1^n, x_2^n)\} \\
&\leq \sum_{(x_1^n, x_2^n)} p(x_1^n, x_2^n) \sum_{\substack{\tilde{x}_2^n \in \mathcal{T}_\epsilon^{(n)}(X_2|x_1^n) \\ \tilde{x}_2^n \neq x_2^n}} \mathsf{P}\{\tilde{x}_2^n \in \mathcal{B}(1)\} \\
&\leq 2^{n(H(X_2|X_1)+\delta(\epsilon))} 2^{-nR_2}
\end{aligned}$$

- Hence, $\mathsf{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R_2 > H(X_2|X_1) + \delta(\epsilon)$
- Remark: Achievability can be proved without time-sharing (NIT 10.3.2)

Extension to more than two sources

Theorem 10.3

The optimal rate region $\mathscr{R}^*(X_1, \ldots, X_k)$ for the k-DMS $(X_1, \ldots, X_k)$ is the set of $(R_1, \ldots, R_k)$ such that

$$\sum_{j \in \mathcal{S}} R_j \geq H(X(\mathcal{S}) | X(\mathcal{S}^c)) \quad \text{for all } \mathcal{S} \subseteq [1:k]$$

- For $k = 3$, $\mathscr{R}^*(X_1, X_2, X_3)$ is the set of (R_1, R_2, R_3) such that

$$\begin{aligned}
R_1 &\geq H(X_1 | X_2, X_3), \\
R_2 &\geq H(X_2 | X_1, X_3), \\
R_3 &\geq H(X_3 | X_1, X_2), \\
R_1 + R_2 &\geq H(X_1, X_2 | X_3), \\
R_1 + R_3 &\geq H(X_1, X_3 | X_2), \\
R_2 + R_3 &\geq H(X_2, X_3 | X_1), \\
R_1 + R_2 + R_3 &\geq H(X_1, X_2, X_3)
\end{aligned}$$

Summary

- k-Component discrete memoryless source (k-DMS)
- Distributed lossless source coding for a k-DMS:
  - Slepian–Wolf optimal rate region
  - Random binning
- Time sharing

References

Cover, T. M. (1975). A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Trans. Inf. Theory*, 21(2), 226–228.

Slepian, D. and Wolf, J. K. (1973). Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4), 471–480.