Encoding of Correlated Sources

GUO Yuanxin, YANG Yufeng, FAN Gang, LIU Zhuoyang

CUHK-Shenzhen

December 9, 2019

Recap: Source Coding

- Recall the data compression part of this course.
- To encode source X, we have shown that a rate $R > H(X)$ is sufficient.
- Similarly, if we have two sources X, Y, a rate of $R > H(X, Y)$ is sufficient if we encode them together.

Separate Encoding: A Schematic Figure

- Now we consider the case when the sources X and Y must be separately described.

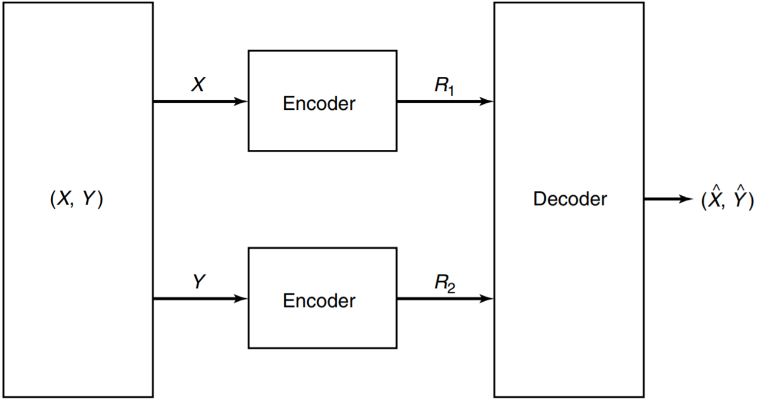


Figure: Separately describing the sources.

Separate Encoding: Rate

- It is obvious that a rate $R = R_x + R_y > H(X) + H(Y)$ is sufficient for separate encoding.
- Intuitively, this should not be the optimal rate, since this rate is sufficient for all possible (X, Y), and it does not utilize the correlation between the sources.
- Surprisingly, it is proved that a total rate $R = H(X, Y)$ is sufficient for separate encoding.

Setup

Definition

A $((2^{nR_1}, 2^{nR_2}), n)$ *distributed source code* for the joint source (X, Y) consists of two encoder maps,

$$f_1 : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR_1}\},$$
$$f_2 : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR_2}\},$$

and a decoder map:

$$g : \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\} \to \mathcal{X}^n \times \mathcal{Y}^n.$$

For a particular sequence pair $(x^n, y^n) \in (\mathcal{X}^n, \mathcal{Y}^n)$, $f_1(x^n), f_2(y^n)$ are the corresponding indices, and (R_1, R_2) is the rate pair of this code.

Setup

Definition

The *probability of error* for a distributed source code is defined as

$$P_e^{(n)} = \mathbb{P}(g(f_1(X^n), f_2(Y^n)) \neq (X^n, Y^n))$$

In other words, an error occur if the recovered sequences are not identical with the original sequences.

Setup

Definition

A rate pair (R_1, R_2) is said to be *achievable* for a distributed source if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ distributed source codes with probability of error $P_e^{(n)} \to 0$. The *achievable rate region* is the closure of the set of achievable rates.

Distributed Source Coding Theorem

Theorem (Slepian-Wolf)

For the distributed source coding problem for the source (X, Y) drawn i.i.d$\sim p(x, y)$, the achievable rate region is given by

$$
\begin{aligned}
R_1 &\geq H(X|Y), \\
R_2 &\geq H(Y|X), \\
R_1 + R_2 &\geq H(X, Y).
\end{aligned}
$$

Figurative Description of Slepian-Wolf Theorem

The achievable rate region described in the Slepian-Wolf Theorem is illustrated below:

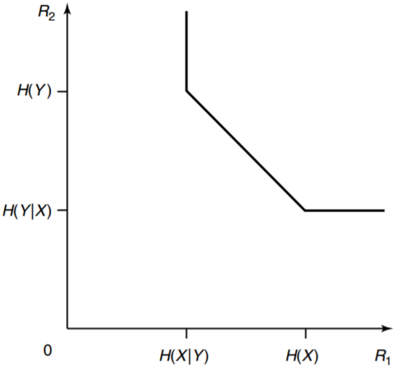


Figure: Rate region for Slepian-Wolf encoding (top-right corner)

Remark & Outline

- The result of Slepian & Wolf is remarkable in that they made use of the correlation of the two sources, while the encoding of the sources are totally independent.
- The following parts will respectively cover the proof of achievability, the proof of the converse, and interpretation with examples.

Achievability of Slepian-Wolf Theorem

Yufeng Yang
117010351

Basic Concept: Coding Procedure Using Random Bins

- Choose index randomly for each source sequence from a sufficiently large set.
- By the property of typical sequences, usually the size of the set formed by typical sequences is small.
- (with high probability) different sequences have different indices. Thus, we can recover the source sequence from the index accurately.

Constructing procedure: Throwing X^n into a row of bins.

- Draw an index random from $\{1,2,3 \ldots \ldots 2^{nR}\}$ for each sequence X^n.
- The sequences X^n & Y^n which have the same index form a bin.
- Look for the typical X^n in the bin. If there is one & only one typical sequence, this will be the estimator $\hat{X}^n$. Otherwise, the error occurred.

Error Analysis of X^n: P_e will go to 0 as n → ∞

- Illustration:

1. If the sequence X^n in the bin is non-typical, there will always be an error.
2. If the set has more than 1 typical sequence, there will be an error.
3. 15.164-15.165 follows from AEP.
4. 15.165-15.166 holds if the rate $R > H(X) + \epsilon$.

$$
\begin{aligned}
P(g(f(\mathbf{X})) \neq \mathbf{X}) &\leq P(\mathbf{X} \notin A_\epsilon^{(n)}) + \sum_{\mathbf{x}} P(\exists \mathbf{x}' \neq \mathbf{x} : \mathbf{x}' \in A_\epsilon^{(n)}, f(\mathbf{x}') \\
&= f(\mathbf{x}))p(\mathbf{x}) \\
&\leq \epsilon + \sum_{\mathbf{x}} \sum_{\substack{\mathbf{x}' \in A_\epsilon^{(n)} \\ \mathbf{x}' \neq \mathbf{x}}} P(f(\mathbf{x}') = f(\mathbf{x}))p(\mathbf{x}) && (15.161) \\
&\leq \epsilon + \sum_{\mathbf{x}} \sum_{\mathbf{x}' \in A_\epsilon^{(n)}} 2^{-nR} p(\mathbf{x}) && (15.162) \\
&= \epsilon + \sum_{\mathbf{x}' \in A_\epsilon^{(n)}} 2^{-nR} \sum_{\mathbf{x}} p(\mathbf{x}) && (15.163) \\
&\leq \epsilon + \sum_{\mathbf{x}' \in A_\epsilon^{(n)}} 2^{-nR} && (15.164) \\
&\leq \epsilon + 2^{n(H(X)+\epsilon)} 2^{-nR} && (15.165) \\
&\leq 2\epsilon && (15.166)
\end{aligned}
$$

Achievability of Slepian-Wolf Theorem. (Set up)

1. **Random code generation:** Assign every sequence X^n to one of 2^{nR_1} bins according to uniform distribution on $\{1,2, \ldots \ldots . 2^{nR_1}\}$. Similarly, assign Y_n to one of 2^{nR_2} bins according to uniform distribution.

2. **Encoding:** Sender 1 sends the index of the bin to which X belongs. Sender 2 sends the index of the bin to which Y belongs.

3. **Decoding:** Given the received pair (i, j), declare $(\hat{x}, \hat{y})$=(x, y) when there is one and only one pair of sequence (**x**, **y**) such that f_1(**x**) = i & f_2(**y**) = j and (**x**, **y**)$\in A_\epsilon$. Otherwise, declare an error.

Achievability of Slepian-Wolf Theorem (Proof)

- One Useful Lemma:

Theorem 15.2.2 *Let S_1, S_2 be two subsets of $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$. For any $\epsilon > 0$, define $A_\epsilon^{(n)}(S_1|\mathbf{s}_2)$ to be the set of $\mathbf{s}_1$ sequences that are jointly ϵ-typical with a particular $\mathbf{s}_2$ sequence. If $\mathbf{s}_2 \in A_\epsilon^{(n)}(S_2)$, then for sufficiently large n, we have*

$$|A_\epsilon^{(n)}(S_1|\mathbf{s}_2)| \le 2^{n(H(S_1|S_2)+2\epsilon)} \tag{15.38}$$

and

$$(1-\epsilon)2^{n(H(S_1|S_2)-2\epsilon)} \le \sum_{\mathbf{s}_2} p(\mathbf{s}_2)|A_\epsilon^{(n)}(S_1|\mathbf{s}_2)|. \tag{15.39}$$

Proof Continued

- 4 Types of error.

1. $E_0=\{(\mathbf{X},\mathbf{Y})\notin A_\epsilon^{(n)}\}$; 2. $E_1=\{\exists\, x' \neq X: f_1(\mathrm{x}') = f_1(\mathrm{x}) \text{ and } (\mathrm{x}',\mathrm{Y}) \in A_\epsilon^{(n)}\}$
3. $E_2=\{\exists\, y' \neq Y: f_2(\mathrm{y}') = f_2(\mathrm{y}) \text{ and } (\mathrm{X},\mathrm{y}') \in A_\epsilon^{(n)}\}$
4. $E_{12}=\{\exists\, (x',y') \neq (X,Y): f_1(\mathrm{x}') = f_1(\mathrm{x}), f_2(\mathrm{y}') = f_2(\mathrm{y}) \text{ and } (\mathrm{x}',\mathrm{y}') \in A_\epsilon^{(n)}\}$

- Formulation:

$$P_\epsilon^{(n)} = \mathrm{P}(E_0 \cup E_1 \cup E_2 \cup E_{12})$$
$$\leq P(E_0) + P(E_1) + P(E_2) + P(E_{12})$$

By property of joint typicality:

$$P(E_0) < \varepsilon \; when\; n\; is\; large.$$

Thus, we focus on E_1, E_2 and E_{12}. (Due to Time limitation, only E_1 will be proved here, the rest are similar.)

Proof Continued

$$P(E_1) = P\{\exists \mathbf{x}' \neq \mathbf{X} : f_1(\mathbf{x}') = f_1(\mathbf{X}), \text{ and } (\mathbf{x}', \mathbf{Y}) \in A_\epsilon^{(n)}\} \quad (15.173)$$

$$= \sum_{(\mathbf{x},\mathbf{y})} p(\mathbf{x}, \mathbf{y}) P\{\exists \mathbf{x}' \neq \mathbf{x} : f_1(\mathbf{x}') = f_1(\mathbf{x}), \ (\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}\} \quad (15.174)$$

$$\leq \sum_{(\mathbf{x},\mathbf{y})} p(\mathbf{x}, \mathbf{y}) \sum_{\substack{\mathbf{x}' \neq \mathbf{x} \\ (\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}}} P(f_1(\mathbf{x}') = f_1(\mathbf{x})) \quad (15.175)$$

$$= \sum_{(\mathbf{x},\mathbf{y})} p(\mathbf{x}, \mathbf{y}) 2^{-nR_1} |A_\epsilon(X|\mathbf{y})| \quad (15.176)$$

$$\leq 2^{-nR_1} 2^{n(H(X|Y)+\epsilon)} \quad \text{(by Theorem 15.2.2),} \quad (15.177)$$

15.174-15.175 follows from the extension of existence of $x'$to the whole set of x'such that $x' \neq x$

15.175-15.176 follows from uniform distribution : $P\big(f_1(x') = f_1(x)\big) = 2^{-nR}$.

15.176-15.177 follows from the Lemma in the last slide.

Proof holds When $R_1 \geq H(X|Y)$. Similarly for E_2, E_{12}. Thus, $R_2 \geq H(X|Y), R_1 + R_2 \geq H(X,Y)$.

Note: All the pictures in the note are from C&T Elements of Information theory.

The Converse for the Slepian-Wolf Theorem

FAN, Gang

Statement:

Statement:

$(X_i, Y_i) \quad \sim \quad p(x, y) \quad i.i.d.$

(R_1, R_2): an achievable rate pair.

Statement:

$(X_i, Y_i) \quad \sim \quad p(x, y) \quad i.i.d.$

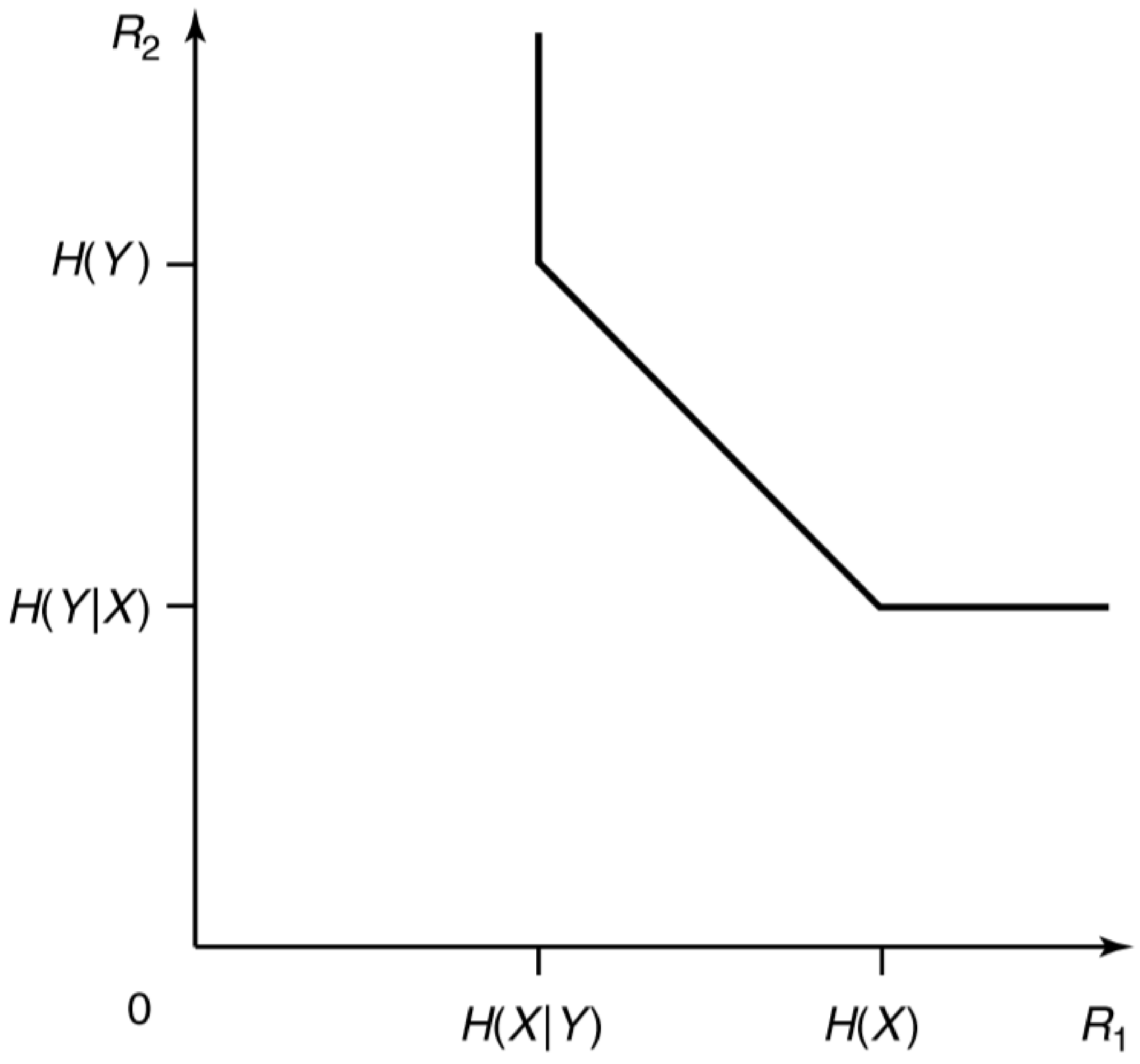
(R_1, R_2): an achievable rate pair.

$$\Rightarrow$$

$$(R_1 + R_2) \geq H(X, Y)$$

$$R_1 \geq H(X|Y)$$

$$R_2 \geq H(Y|X)$$



R_2
$H(Y)$
$H(Y|X)$
0
$H(X|Y)$
$H(X)$
R_1

Proof:

Proof:

Since (R_1, R_2) is achievable, there exists a sequence of

$$\left((2^{nR_1}, 2^{nR_2}), n\right)$$

distributed source codes that $P_e^{(n)} \to 0$ as $n \to \infty$.

Proof:

Since (R_1, R_2) is achievable, there exists a sequence of

$$\left((2^{nR_1}, 2^{nR_2}), n\right)$$

distributed source codes that $P_e^{(n)} \to 0$ as $n \to \infty$.

Define:

Encoding functions f_1, f_2

Decoding function g

$$I_0 \coloneqq f_1(X^n)$$

$$J_0 \coloneqq f_2(Y^n)$$

Proof:

$$\Rightarrow (R_1 + R_2) \geq H(X,Y)$$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$$\Rightarrow (R_1 + R_2) \geq H(X, Y)$$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$= I(X^n, Y^n; I_0, J_0)$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$= I(X^n, Y^n; I_0, J_0)$

$= H(X^n, Y^n) - H(X^n, Y^n | I_0, J_0)$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof: (Fano's inequality)

$$H(X^n, Y^n | I_0, J_0)$$

$$\leq H_B\left(P_e^{(n)}\right) + P_e^{(n)} \log(|\mathcal{X}^n \times \mathcal{Y}^n| - 1)$$

Proof: (Fano's inequality)

$$H(X^n, Y^n | I_0, J_0)$$

$$\leq H_B\left(P_e^{(n)}\right) + P_e^{(n)} \log(|\mathcal{X}^n \times \mathcal{Y}^n| - 1)$$

$$\leq 1 + nP_e^{(n)}(\log|\mathcal{X}| + \log|\mathcal{Y}|)$$

Proof: (Fano's inequality)

$H(X^n, Y^n | I_0, J_0)$

$\leq H_B\left(P_e^{(n)}\right) + P_e^{(n)} \log(|\mathcal{X}^n \times \mathcal{Y}^n| - 1)$

$\leq 1 + nP_e^{(n)}(\log|\mathcal{X}| + \log|\mathcal{Y}|)$

$\leq 1 + n\varepsilon_n$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$= I(X^n, Y^n; I_0, J_0)$

$= H(X^n, Y^n) - H(X^n, Y^n | I_0, J_0)$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$= I(X^n, Y^n; I_0, J_0)$

$= H(X^n, Y^n) - H(X^n, Y^n | I_0, J_0)$

$\geq H(X^n, Y^n) - n\varepsilon_n$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof:

$n(R_1 + R_2)$

$\geq H(I_0, J_0)$

$= I(X^n, Y^n; I_0, J_0) + H(I_0, J_0 | X^n, Y^n)$

$= I(X^n, Y^n; I_0, J_0)$

$= H(X^n, Y^n) - H(X^n, Y^n | I_0, J_0)$

$\geq H(X^n, Y^n) - n\varepsilon_n$

$= nH(X, Y) - n\varepsilon_n$

$\Rightarrow (R_1 + R_2) \geq H(X, Y)$

Proof:

nR_1
$\geq H(I_0)$
$\geq H(I_0|Y^n)$
$= I(X^n; I_0|Y^n) + H(I_0|X^n, Y^n)$
$= I(X^n; I_0|Y^n)$
$= H(X^n|Y^n) - H(X^n|I_0, Y_n)$
$\geq H(X^n|Y^n) - H(X^n|I_0, J_0, Y_n)$
$\geq H(X^n|Y^n) - n\varepsilon_n$
$= nH(X|Y) - n\varepsilon_n$

$\Rightarrow R_1 \geq H(X|Y)$

Statement:

$(X_i, Y_i) \quad \sim \quad p(x, y) \quad i.i.d.$

(R_1, R_2): an achievable rate pair.

$$\Rightarrow$$

$$(R_1 + R_2) \geq H(X, Y)$$

$$R_1 \geq H(X|Y)$$

$$R_2 \geq H(Y|X)$$

Many-Source Case:

Many-Source Case:

$(X_{1i}, X_{2i}, \dots, X_{mi}) \quad \sim \quad p(x_1, x_2, \dots, x_m) \quad i.i.d.$

$(R_1, \dots, R_m)$: an achievable rate vector

Many-Source Case:

$(X_{1i}, X_{2i}, \dots, X_{mi}) \quad \sim \quad p(x_1, x_2, \dots, x_m) \quad i.i.d.$

$(R_1, \dots, R_m)$: an achievable rate vector

$$\Rightarrow$$

$$\forall S \subseteq \{1,2,\dots,m\}: R(S) \coloneqq \sum_{i \in S} R_i > H(X(S)|X(S^C))$$

$$(X(S) \coloneqq \{X_i : i \in S\})$$

Proof: (Many-Source Case)

$$n\sum_{i\in S} R_i \geq H(I_0) \geq H\left(I_0|X^n(S^C)\right)$$

$$= I\left(X^n(S); I_0|X^n(S^C)\right) + H\left(I_0|X^n(S), X^n(S^C)\right)$$

$$= I\left(X^n(S); I_0|X^n(S^C)\right)$$

$$= H\left(X^n(S)|X^n(S^C)\right) - H\left(X^n(S)|I_0, X^n(S^C)\right)$$

$$\geq H\left(X^n(S)|X^n(S^C)\right) - H\left(X^n(S)|I_0, J_0, X^n(S^C)\right)$$

$$> H\left(X^n(S)|X^n(S^C)\right) - n\varepsilon_n$$

$$= nH\left(X^n(S)|X^n(S^C)\right) - n\varepsilon_n$$

$$\Rightarrow \sum_{i\in S} R_i > H(X(S)|X(S^C))$$

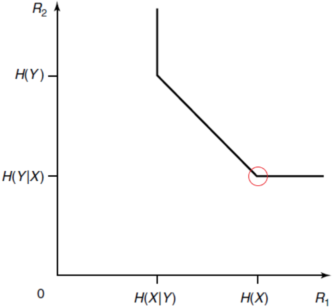
Interpretation

Execution

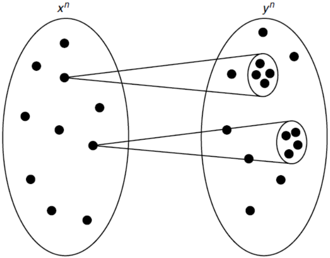
Interpretation

Consider the corner point with rate $R_1 = H(X), R_2 = H(Y|X)$.

- Can encode X^n efficiently with $nH(X)$ bits
- Want to encode Y^n without knowing X^n with $nH(X|Y)$ bits



(a) Rate region



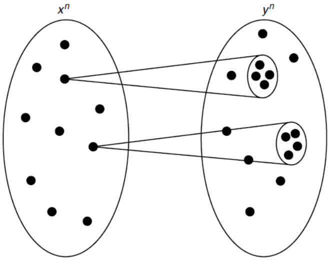
(b) Jointly typical fans

Joint coding case Knows X^n; send index of Y^n within typical fan
Distributed case NOT know X^n; send the color of Y^n randomly assigned, 2^{nR_2} colors, $R_2 > H(Y|X) \to$ single instance in one fan

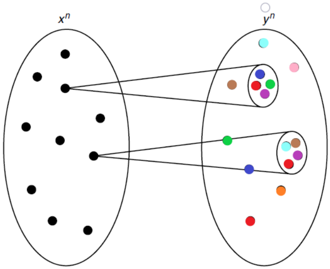
Interpretation

Consider the corner point with rate $R_1 = H(X), R_2 = H(Y|X)$.

- Can encode X^n efficiently with $nH(X)$ bits
- Want to encode Y^n without knowing X^n with $nH(X|Y)$ bits



(c) Rate region



(d) Jointly typical fans

Joint coding case Knows X^n; send index of Y^n within typical fan
Distributed case NOT know X^n; send the color of Y^n randomly assigned, 2^{nR_2} colors, $R_2 > H(Y|X) \to$ single instance in one fan

Execution: A concrete example

$X \sim Ber(\frac{1}{2})$, $Y \sim Ber(\frac{1}{2})$, X, Y correlated with $Pr(X \neq Y) = p$.
Consider compression of (x^n, y^n) using Slepian-Wolf coding (rate $1 + H(p)$)

- X compresses up to $H(X) = 1$ bit $\Rightarrow x^n$ compresses in n bits
- Y compresses in the following way (for y^n)
  - Take random Bernoulli matrix $\Phi_{m \times n}$, $m < n$
  - Compute $\Phi y^n \mod 2$ $\Rightarrow$ this gives a m-bits compression for y^n
- On average, this gives a compression rate of $1 + \frac{m}{n}$

Then consider the decoding

- Decode to get x^n
- The task becomes finding y^n
  - Consider $\Phi z^n \mod 2 = \Phi(x^n - y^n) \mod 2$, where $z = (z_1\ z_2 \cdots z_n) = x^n - y^n$ is $i.i.d$ with $Pr(z^i = 1) = p$
  - Want to find out z^n from $\Phi z^n \mod 2$ $\Rightarrow y^n = x^n + z^n$

Now analyze the error rate

- Typical set of z^n: $2^{-n(H(z)+\epsilon)} \leq \left|A_\epsilon^{(n)}(z)\right| \leq 2^{-n(H(z)-\epsilon)}$,
  $Pr(A_\epsilon^{(n)}(z)) > 1 - \epsilon \Rightarrow avgP_e \leq \epsilon + 2^{n(H(z)+\epsilon)}2^{-m} \to 0$