Common Information

Abbas El Gamal

Stanford University

Viterbi Lecture, USC, April 2014

Andrew Viterbi's Fabulous Formula, *IEEE Spectrum*, 2010

Andrew Viterbi's Fabulous Formula, *IEEE Spectrum*, 2010

"A couple of professors at Stanford were being quoted in the press saying that CDMA violated the laws of physics."

About this lecture

- Common information between correlated information sources

About this lecture

- Common information between correlated information sources
- The lecture is tutorial in nature with many examples

About this lecture

- Common information between correlated information sources
- The lecture is tutorial in nature with many examples
- No proofs, but a potentially interesting framework and new result

About this lecture

- Common information between correlated information sources
- The lecture is tutorial in nature with many examples
- No proofs, but a potentially interesting framework and new result

- Outline:
  - Brief introduction to information theory
  - Information sources and measuring information

About this lecture

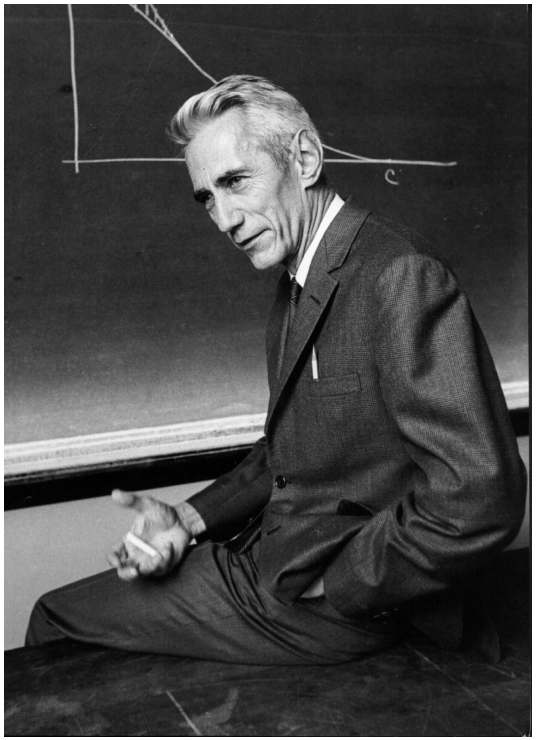
- Common information between correlated information sources
- The lecture is tutorial in nature with many examples
- No proofs, but a potentially interesting framework and new result

- Outline:
  - Brief introduction to information theory
  - Information sources and measuring information
  - Brief introduction to network information theory
  - Correlated information sources and measuring common information

About this lecture

- Common information between correlated information sources
- The lecture is tutorial in nature with many examples
- No proofs, but a potentially interesting framework and new result
- Outline:
  - Brief introduction to information theory
  - Information sources and measuring information
  - Brief introduction to network information theory
  - Correlated information sources and measuring common information
- Some of the basic models, ideas, and results of information theory

Information theory



The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning . . . . These semantic aspects of communication are irrelevant to the engineering problem.

– A Mathematical Theory of Communication, Shannon (1948)

Information theory

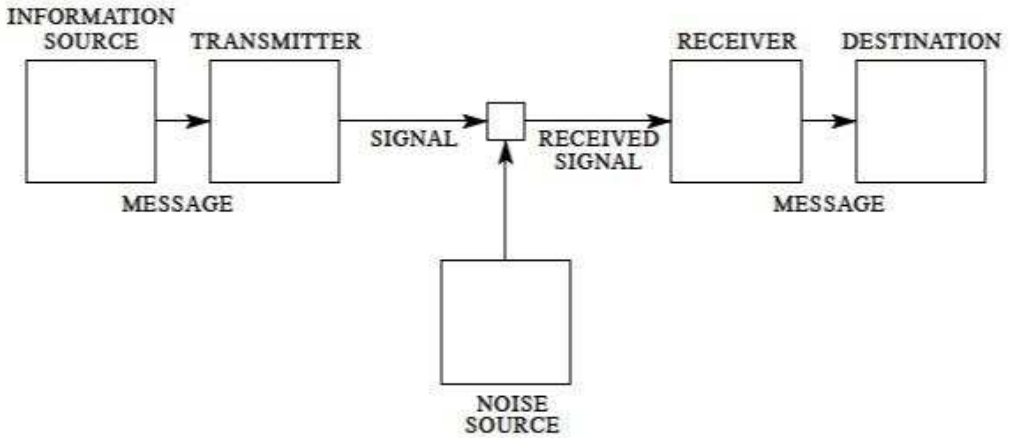


Fig. 1 — Schematic diagram of a general communication system.

- Mathematical models for the source and the channel

Information theory

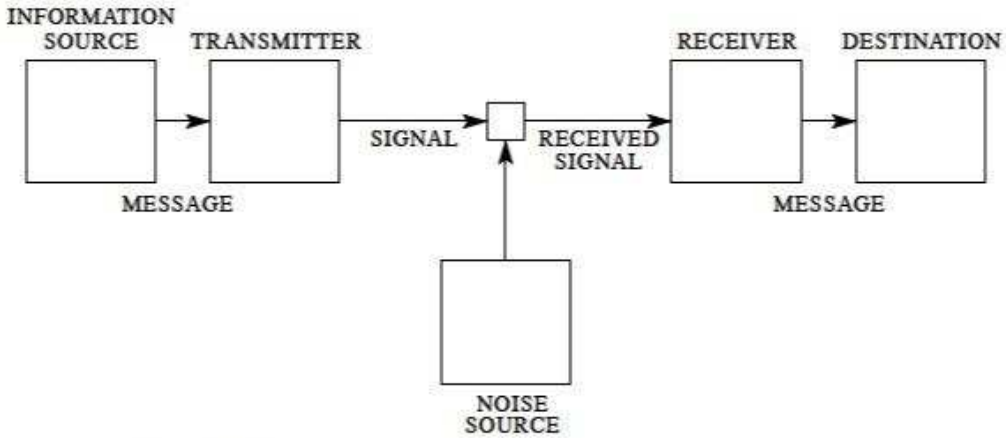


Fig. 1 — Schematic diagram of a general communication system.

- Mathematical models for the source and the channel
- Measures of information (entropy and mutual information)
- Limits on compression and communication

Information theory

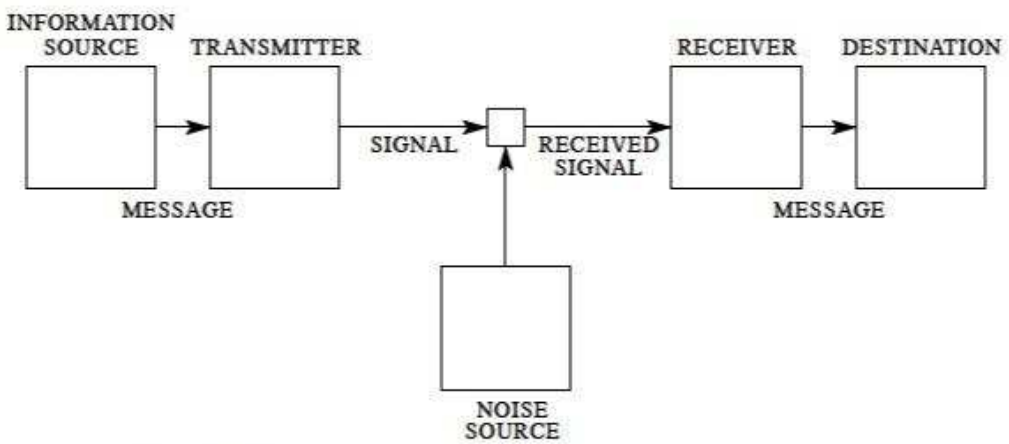


Fig. 1 — Schematic diagram of a general communication system.

- Mathematical models for the source and the channel
- Measures of information (entropy and mutual information)
- Limits on compression and communication
- Bits as a universal interface between the source and the channel

Information theory

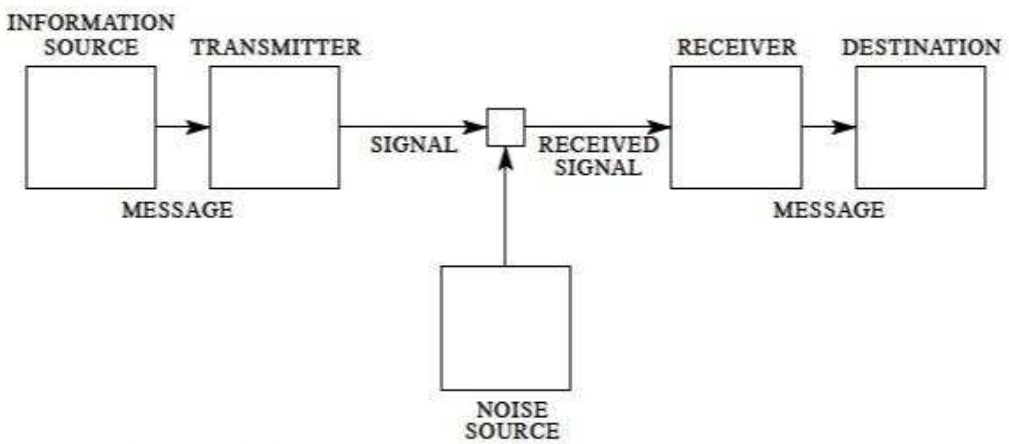
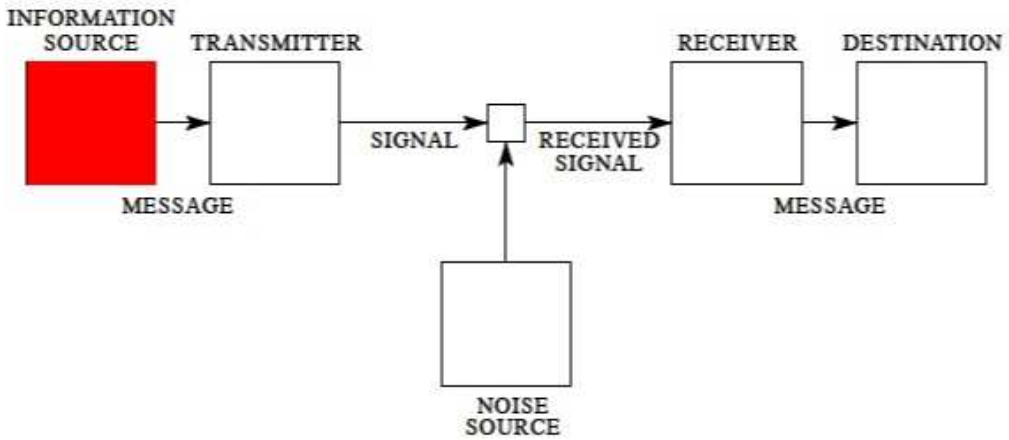


Fig. 1 — Schematic diagram of a general communication system.

- Coding schemes that achieve the Shannon limits
  - Viterbi decoding algorithm (Viterbi 1967)
  - Lempel–Ziv compression algorithm (Ziv–Lempel 1977, 1978)
  - Trellis coded modulation (Ungerboeck 1982)
  - Turbo codes (Berrou–Glavieux 1996)
  - LDPC codes (Gallager 1963, Richardson–Urbanke 2008)
  - Polar codes (Arıkan 2009)

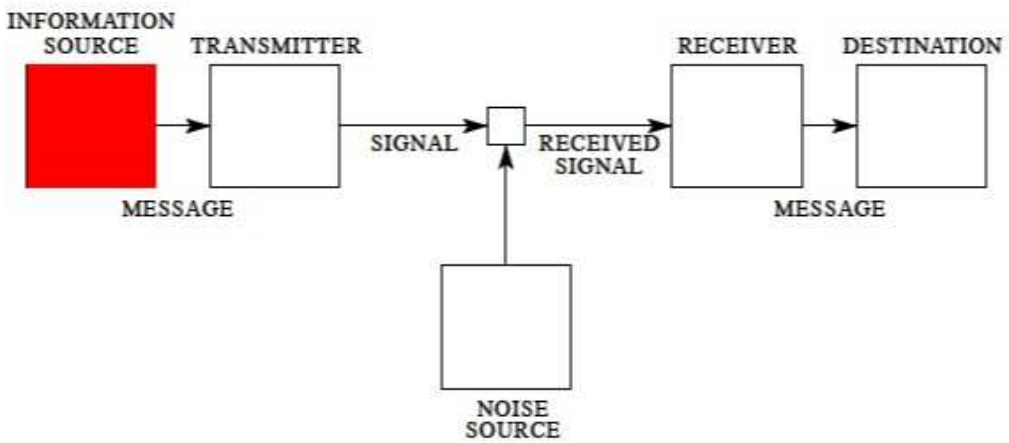
Information source



An information source which produces a message or sequence of messages to be communicated to the receiving terminal. The message may be of various types: e.g. (a) A sequence of letters as in a telegraph of teletype system; (b) A single function of time $f(t)$ as in radio or telephony; . . .

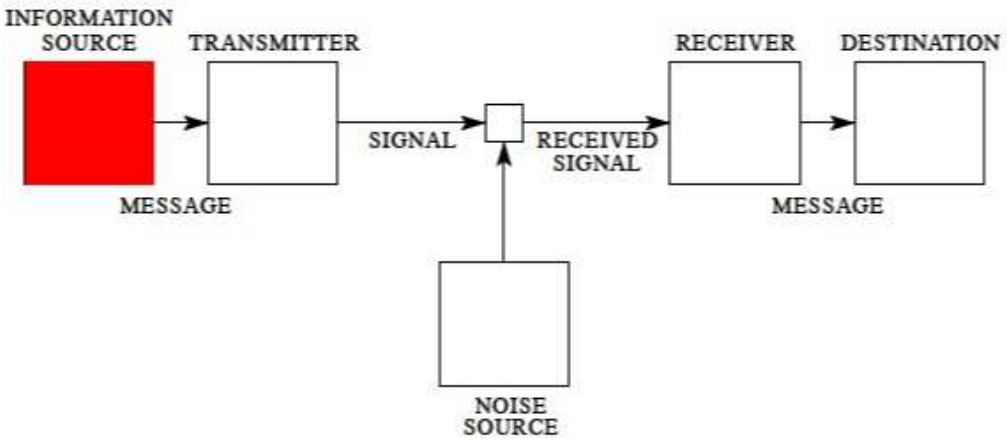
– A Mathematical Theory of Communication, Shannon (1948)

Modeling the information source



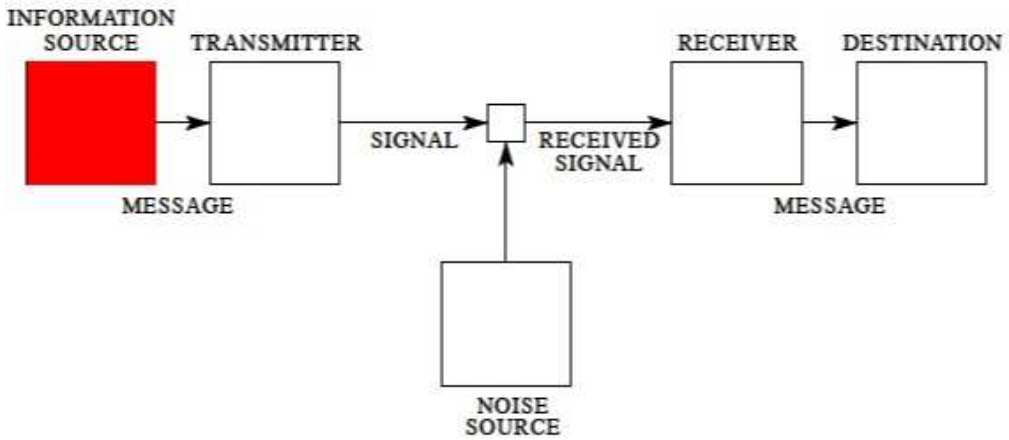
We can think of a discrete source as generating the message, symbol by symbol. . . . A physical system, or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities, is known as a stochastic process.

Measuring information of the source



Can we define a quantity which will measure, in some sense, how much information is "produced" by such a process, or better, at what rate information is produced?

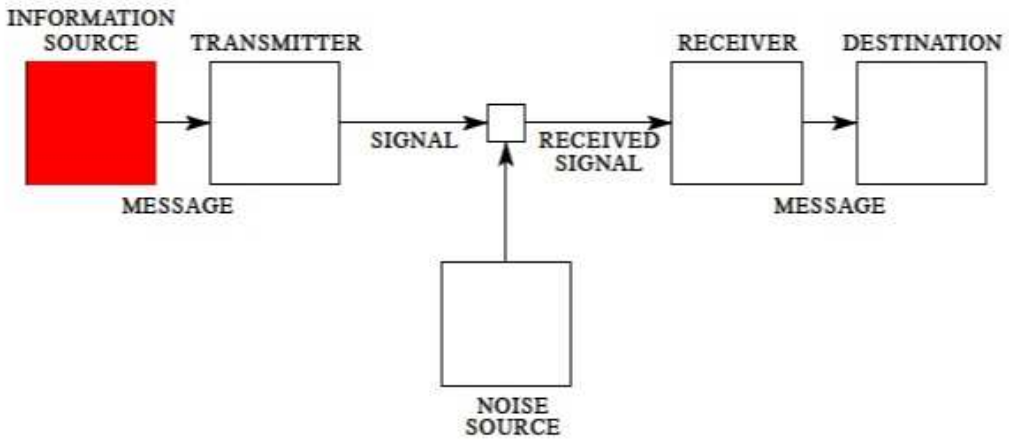
Measuring information of the source



Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. *. . . If there is such a measure, say* $H(p_1, p_2, \ldots, p_n)$, *it is reasonable to require of it the following properties:*

1. H *should be continuous in the* p_i.
2. *If all the* p_i *are equal,* H *should be monotonic increasing function of* n.
3. *If a choice be broken down into two successive choices, the original* H *should be the weighted sum of the individual values of* H.

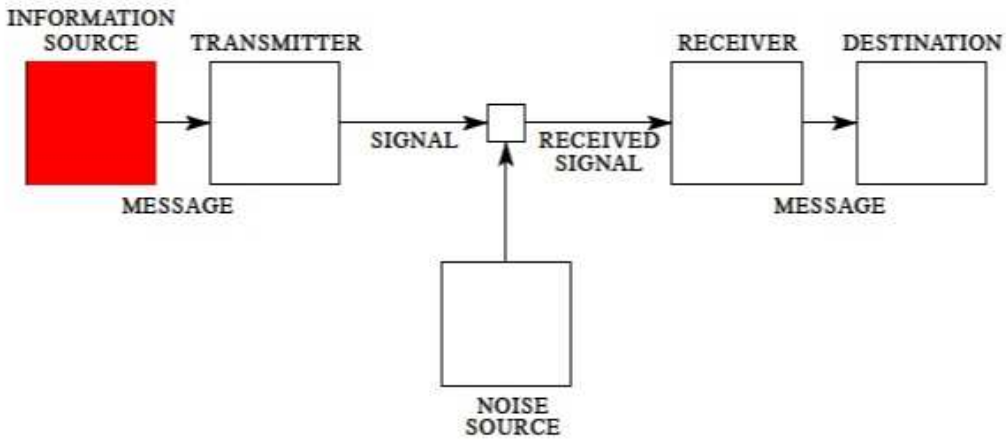
Measuring information of the source



Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

Measuring information of the source



Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

This theorem, and the assumptions required for its proof, are in no way necessary for the present theory. . . . The real justification of these definitions, however, will reside in their implications.

– A Mathematical Theory of Communication, Shannon (1948)

Setups for measuring information

- Entropy arises naturally as a measure of information in many settings

Setups for measuring information

- Entropy arises naturally as a measure of information in many settings
- Will describe three of them
  - Zero-error compression
  - Randomness extraction
  - Source simulation

Setups for measuring information

- Entropy arises naturally as a measure of information in many settings
- Will describe three of them
  - Zero-error compression
  - Randomness extraction
  - Source simulation
- Assume a discrete memoryless source $(\mathcal{X}, p(x))$ (DMS X)

  X generates independent identically distributed (i.i.d.) $X_1, X_2, \ldots \sim p(x)$

Setups for measuring information

- Entropy arises naturally as a measure of information in many settings
- Will describe three of them
  - Zero-error compression
  - Randomness extraction
  - Source simulation
- Assume a discrete memoryless source $(\mathcal{X}, p(x))$ (DMS X)

  X generates independent identically distributed (i.i.d.) $X_1, X_2, \ldots \sim p(x)$
- Example: Bernoulli source with parameter $p \in [0,1]$ (Bern(p) source)

  $\mathcal{X} = \{0,1\}$, $p_X(1) = p$; generates i.i.d $X_1, X_2, \ldots \sim \text{Bern}(p)$

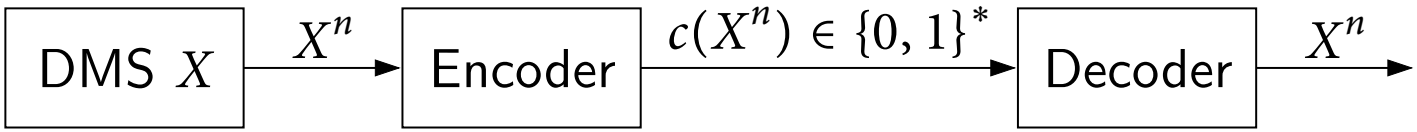
Setups for measuring information

- Entropy arises naturally as a measure of information in many settings
- Will describe three of them
  - Zero-error compression
  - Randomness extraction
  - Source simulation
- Assume a discrete memoryless source $(\mathcal{X}, p(x))$ (DMS X)

  X generates independent identically distributed (i.i.d.) $X_1, X_2, \ldots \sim p(x)$
- Example: Bernoulli source with parameter $p \in [0,1]$ (Bern(p) source)

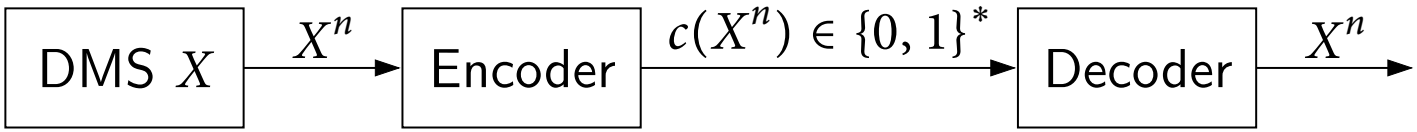
  $\mathcal{X} = \{0,1\}$, $p_X(1) = p$; generates i.i.d $X_1, X_2, \ldots \sim \text{Bern}(p)$
- Example: "DNA source," $\mathcal{X} = \{\text{A},\text{G},\text{C},\text{T}\}$; $p(x) = [p_1 \;\; p_2 \;\; p_3 \;\; p_4]$

Zero-error compression

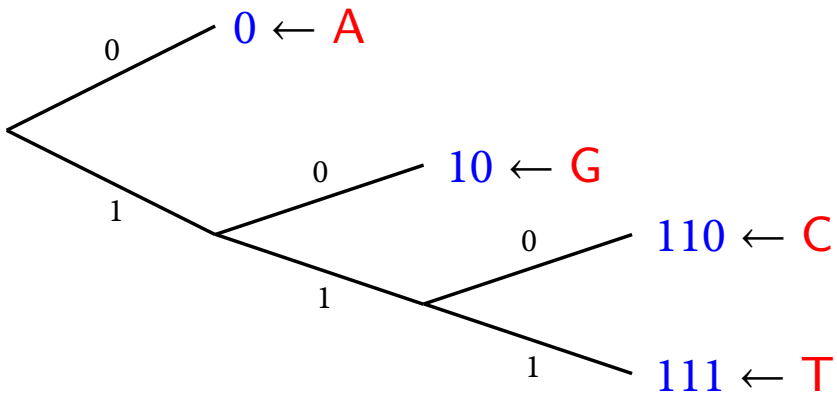


- $X^n = (X_1, X_2, \ldots, X_n)$ i.i.d. $\sim p(x)$
- Variable-length prefix-free code: $c(x^n) \in \{0,1\}^* = \{0, 1, 00, 01, 10, 11, \ldots\}$

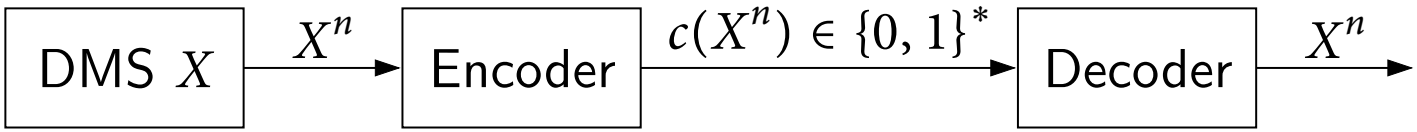
Zero-error compression



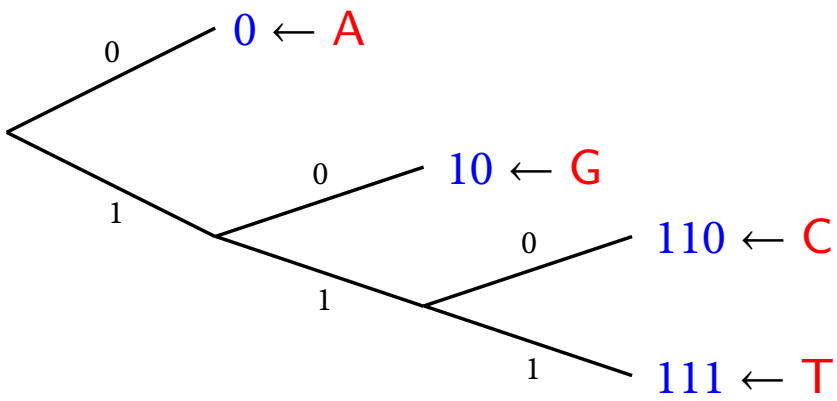
- $X^n = (X_1, X_2, \ldots, X_n)$ i.i.d. $\sim p(x)$
- Variable-length prefix-free code: $c(x^n) \in \{0,1\}^* = \{0, 1, 00, 01, 10, 11, \ldots\}$
- Example: $\mathcal{X} = \{\text{A,G,C,T}\}$, $p(x) = [5/8 \;\; 1/4 \;\; 1/16 \;\; 1/16]$, $n = 1$



Zero-error compression



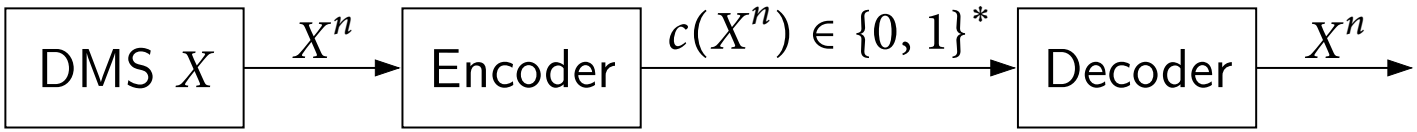
- $X^n = (X_1, X_2, \ldots, X_n)$ i.i.d. $\sim p(x)$
- Variable-length prefix-free code: $c(x^n) \in \{0,1\}^* = \{0, 1, 00, 01, 10, 11, \ldots\}$
- Example: $\mathcal{X} = \{\text{A,G,C,T}\}$, $p(x) = [5/8 \;\; 1/4 \;\; 1/16 \;\; 1/16]$, $n = 1$



- Let L_n be the codeword length and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol

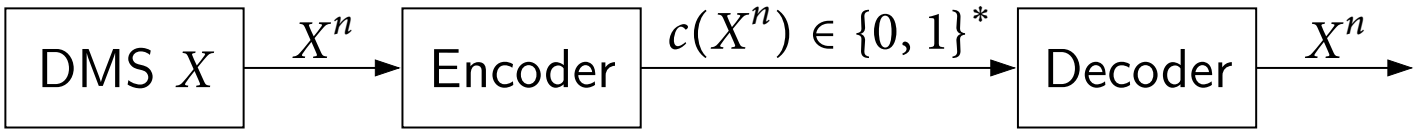
$$\mathsf{E}(L_1) = 1 \cdot \frac{5}{8} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} = 1.5 \text{ bits}$$

Zero-error compression



- $X^n = (X_1, X_2, \ldots, X_n)$ i.i.d. $\sim p(x)$
- Variable-length prefix-free code: $c(x^n) \in \{0,1\}^* = \{0, 1, 00, 01, 10, 11, \ldots\}$
- Let L_n be the codeword length and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol
- Measure the information rate of X by: $R^* = \inf_n \min_{\text{codes}} R_n$

Zero-error compression



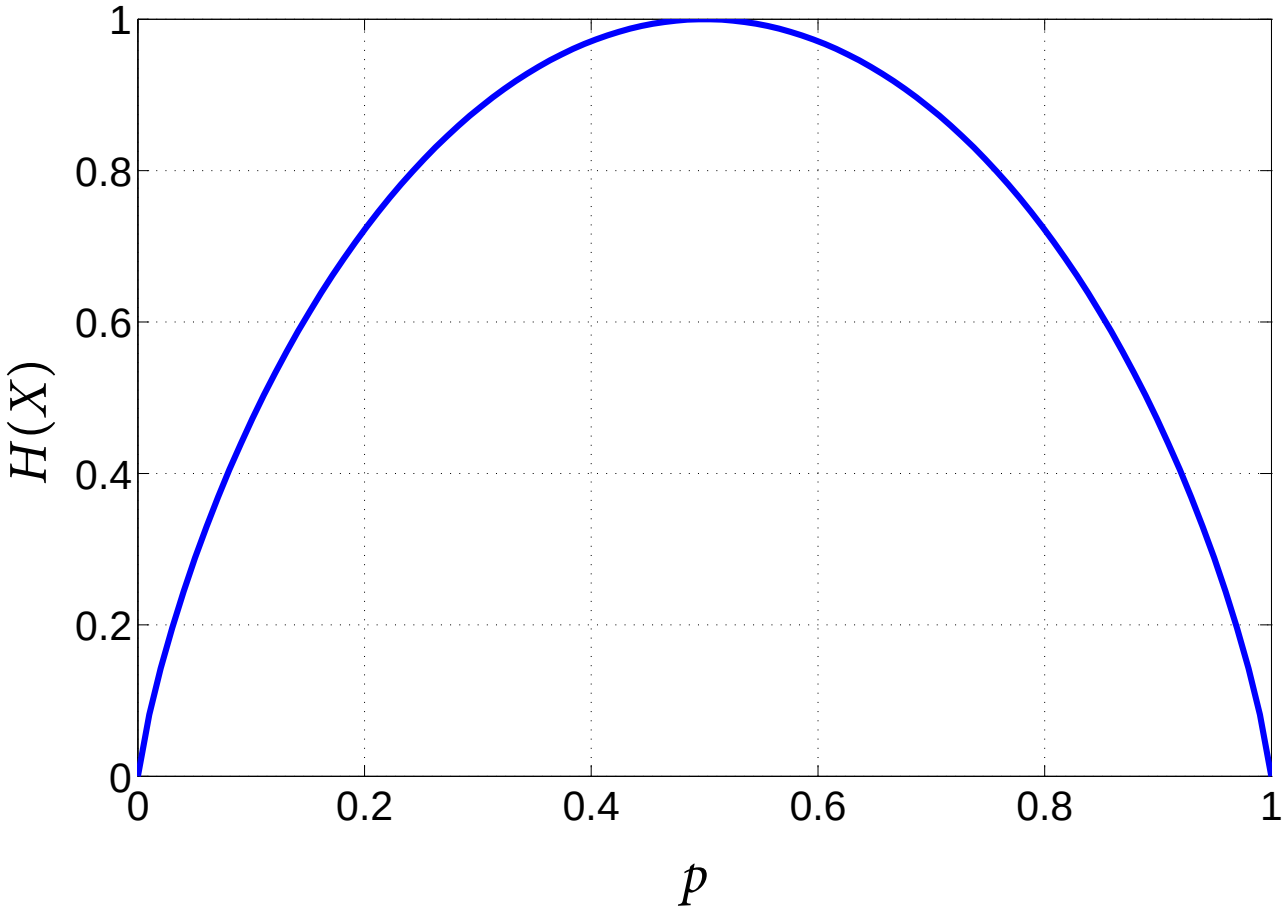
- $X^n = (X_1, X_2, \ldots, X_n)$ i.i.d. $\sim p(x)$
- Variable-length prefix-free code: $c(x^n) \in \{0,1\}^* = \{0, 1, 00, 01, 10, 11, \ldots\}$
- Let L_n be the codeword length and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol
- Measure the information rate of X by: $R^* = \inf_n \min_{\text{codes}} R_n$

Information rate is the entropy (Shannon 1948)

$$R^* = H(X) = \sum_x -p(x) \log p(x) \quad \text{bits/symbol}$$

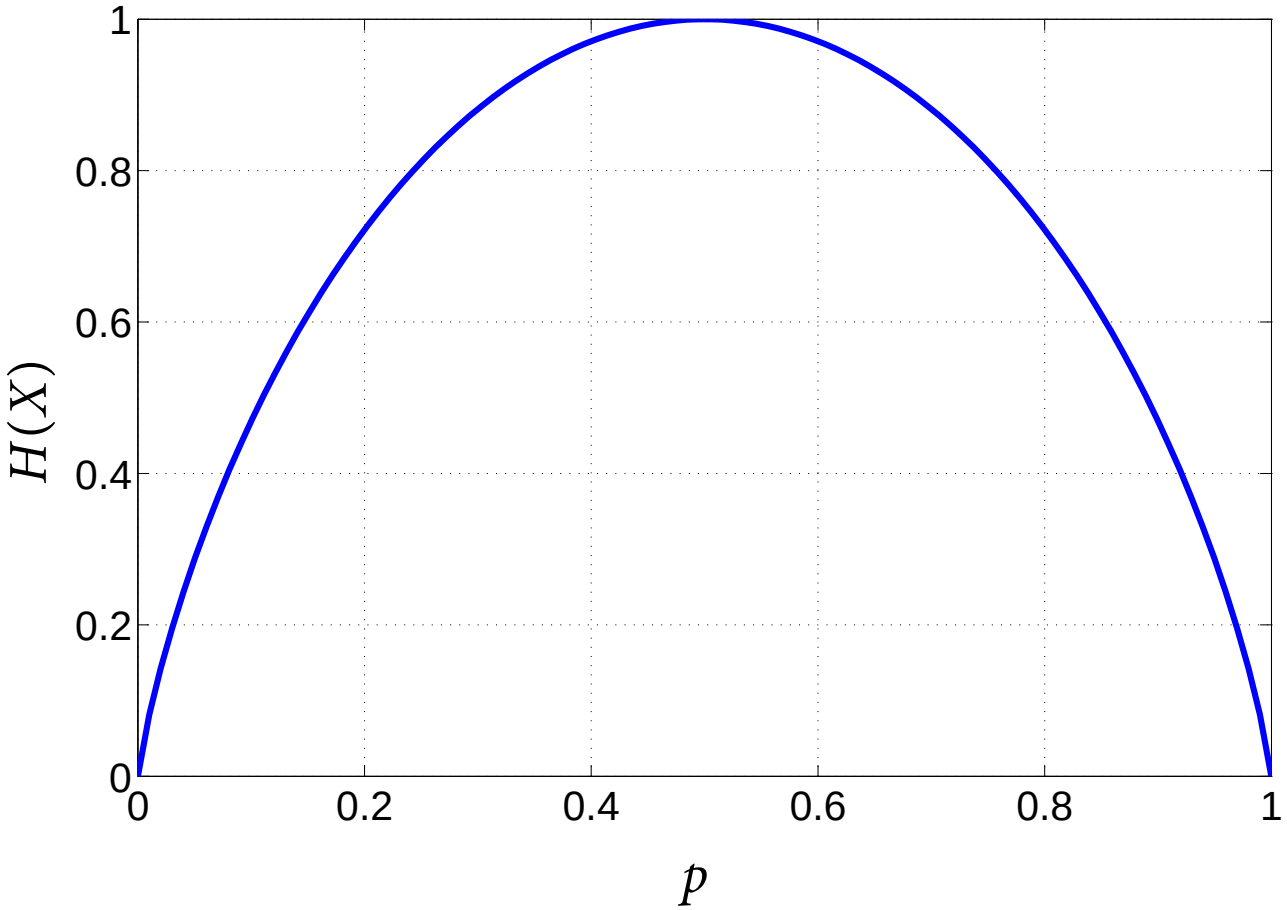
Entropy examples

- $X \sim \text{Bern}(p)$: $H(X) = H(p) = -p \log p - (1-p)\log(1-p)$



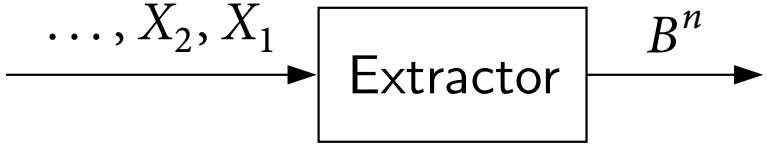
Entropy examples

- $X \sim \text{Bern}(p)$: $H(X) = H(p) = -p \log p - (1-p)\log(1-p)$



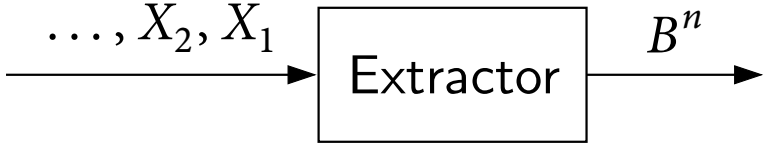
- For DNA source: $H(X) = 1.424$ bits $(\mathsf{E}(L_1) = 1.5)$

Randomness extraction

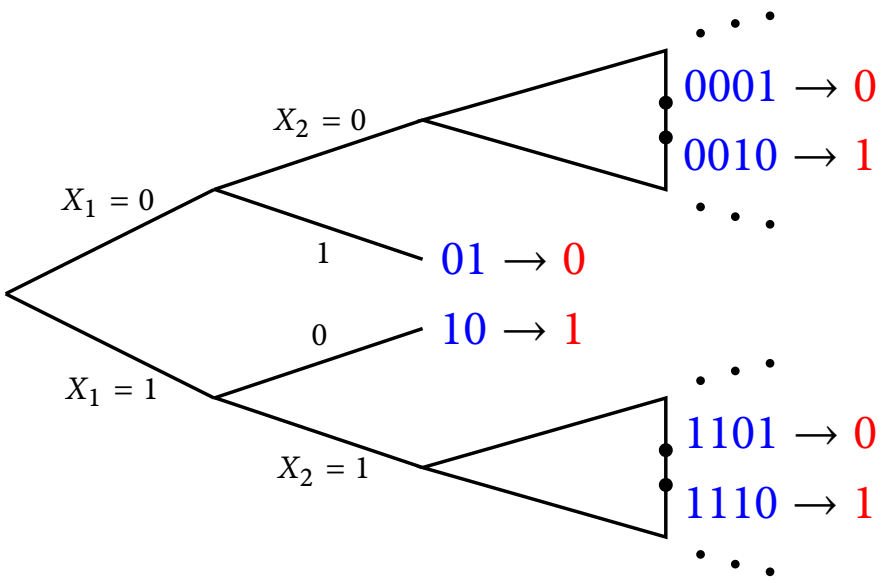


- X is a DMS; B^n is an i.i.d. $\sim \text{Bern}(1/2)$ sequence
- Let L_n be length of X sequence and $R_n = n/\,\mathsf{E}(L_n)$ bits/symbol

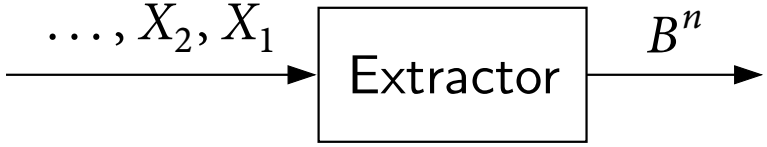
Randomness extraction



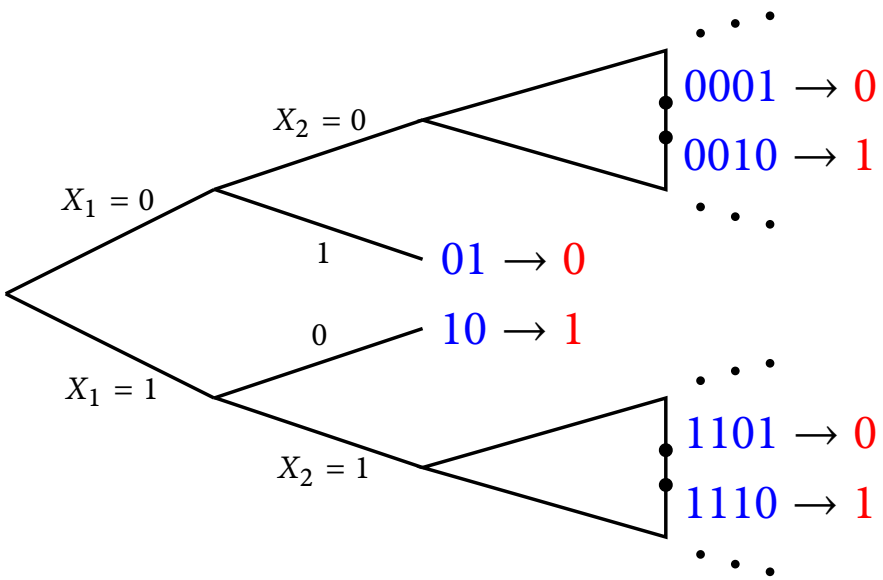
- X is a DMS; B^n is an i.i.d. $\sim$ Bern(1/2) sequence
- Let L_n be length of X sequence and $R_n = n/\,\mathsf{E}(L_n)$ bits/symbol
- Example: Let X be Bern(1/3) source, $n = 1$ $(B_1 \sim \text{Bern}(1/2))$



Randomness extraction

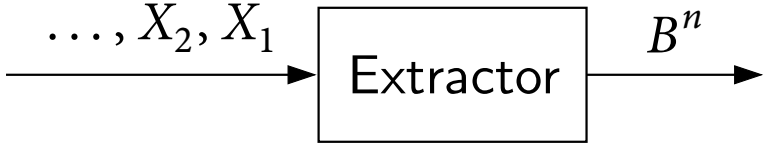


- X is a DMS; B^n is an i.i.d. $\sim$ Bern(1/2) sequence
- Let L_n be length of X sequence and $R_n = n/\,\mathsf{E}(L_n)$ bits/symbol
- Example: Let X be Bern(1/3) source, $n = 1$ $(B_1 \sim \text{Bern}(1/2))$



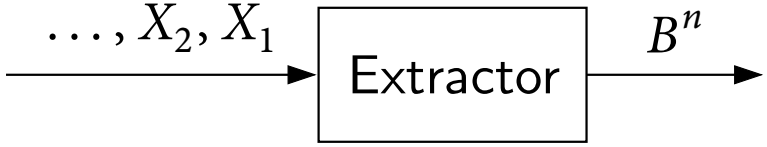
$$\mathsf{E}(L_1) = \frac{4}{9} \cdot 2 + \frac{5}{9} \cdot (2 + \mathsf{E}(L_1)) \;\Rightarrow\; R_1 = 1/\,\mathsf{E}(L_1) = \frac{2}{9}$$

Randomness extraction



- X is a DMS; B^n is an i.i.d. $\sim \mathrm{Bern}(1/2)$ sequence
- Let L_n be length of X sequence and $R_n = n/\,\mathsf{E}(L_n)$ bits/symbol
- The information rate of X: $R^* = \sup_n \max_{\text{extractor}} R_n$

Randomness extraction

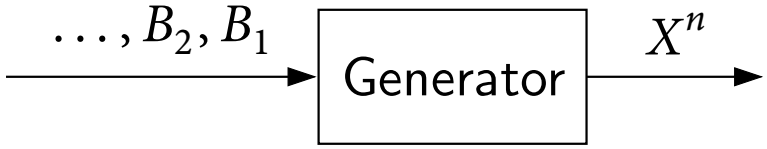


- X is a DMS; B^n is an i.i.d. $\sim \text{Bern}(1/2)$ sequence
- Let L_n be length of X sequence and $R_n = n/\,\mathsf{E}(L_n)$ bits/symbol
- The information rate of X: $R^* = \sup_n \max_{\text{extractor}} R_n$

Information rate is the entropy (Elias 1972)

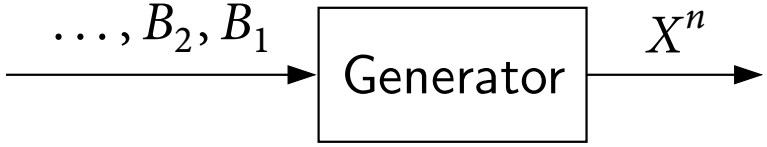
$$R^* = H(X) \quad \text{bits/symbol}$$

Source simulation

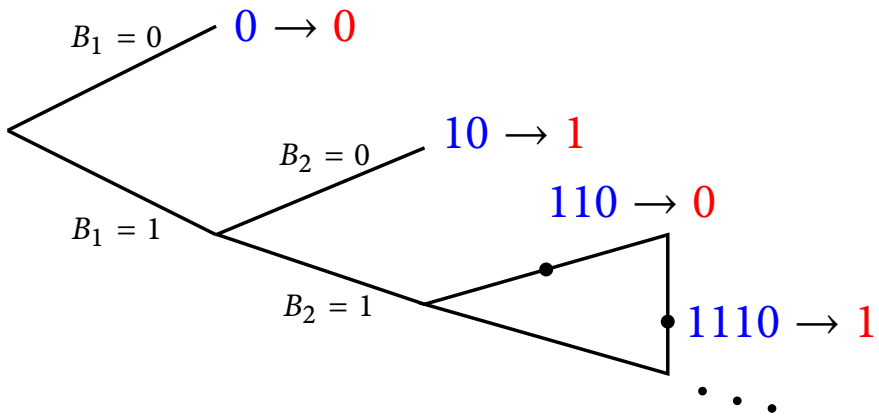


- B is a Bern$(1/2)$ source; X^n i.i.d. $\sim p(x)$ sequence
- Let L_n be length of the B sequence and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol

Source simulation



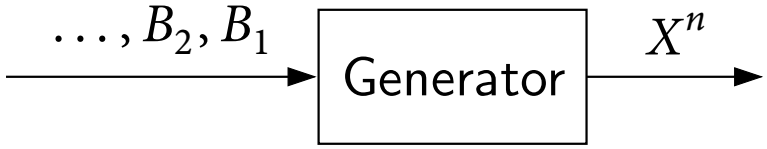
- B is a Bern$(1/2)$ source; X^n i.i.d. $\sim p(x)$ sequence
- Let L_n be length of the B sequence and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol
- Example: Let X be Bern$(1/3)$ source, $n = 1$



$$\mathsf{P}\{X_1 = 1\} = \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{1}{3}$$

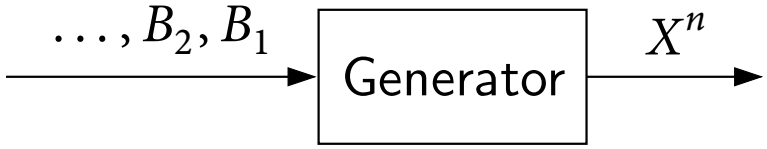
$$\mathsf{E}(L_1) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot (2 + \mathsf{E}(L_1)) \;\Rightarrow\; \mathsf{E}(L_1) = 2$$

Source simulation



- B is a $\mathrm{Bern}(1/2)$ source; X^n i.i.d. $\sim p(x)$ sequence
- Let L_n be length of the B sequence and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol
- The information rate of X: $R^* = \inf_n \min_{\text{generator}} R_n$

Source simulation



- B is a $\mathrm{Bern}(1/2)$ source; X^n i.i.d. $\sim p(x)$ sequence
- Let L_n be length of the B sequence and $R_n = (1/n)\,\mathsf{E}(L_n)$ bits/symbol
- The information rate of X: $R^* = \inf_n \min_{\text{generator}} R_n$

Information rate is the entropy (Knuth–Yao 1976)

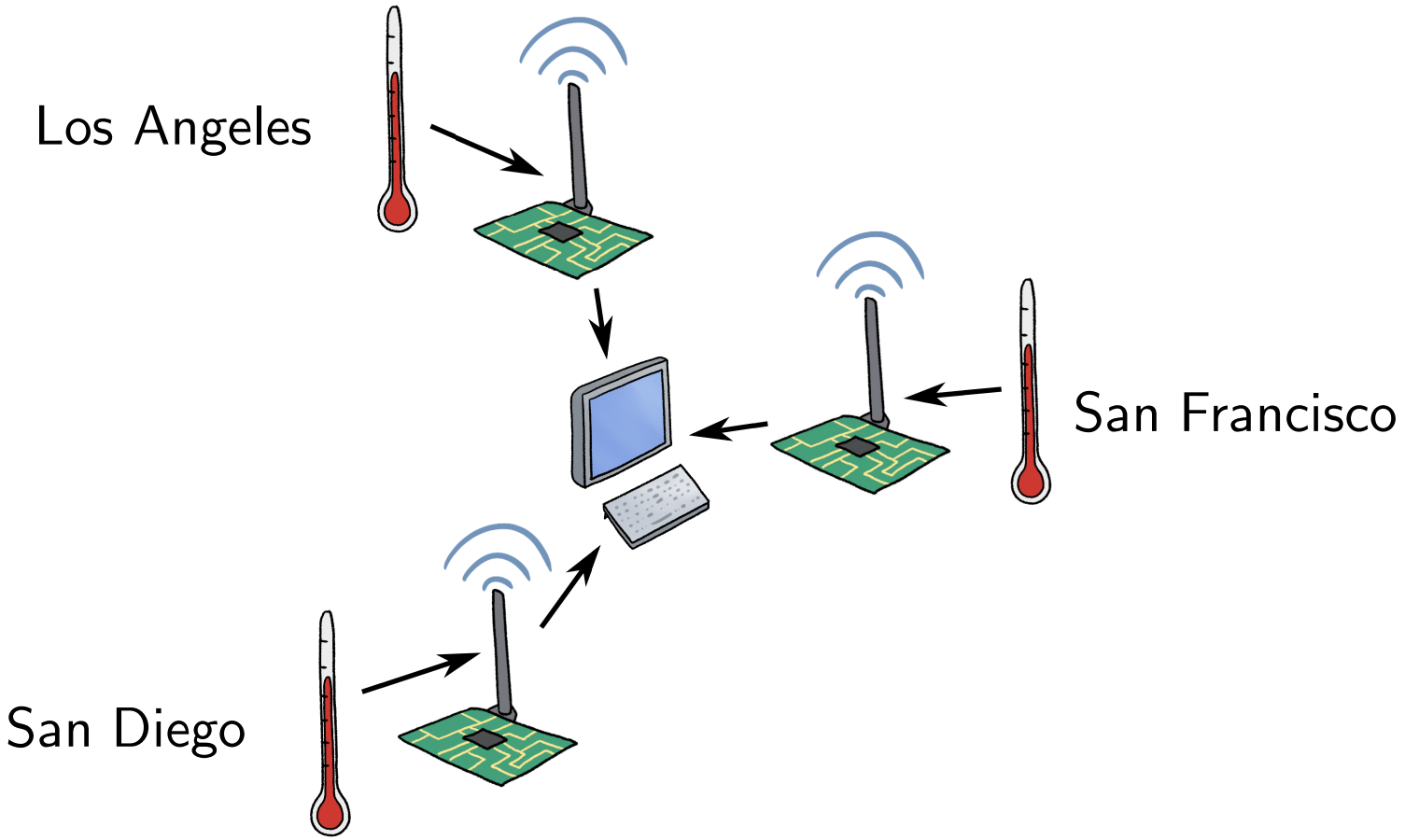
$$R^* = H(X) \quad \text{bits/symbol}$$

Summary

- Entropy arises naturally as a measure of information rate:
  - The minimum average description length in bits/symbol of X
  - The maximum number of bits/symbol that can be extracted from X
  - The minimum number of bits/symbol needed to simulate X

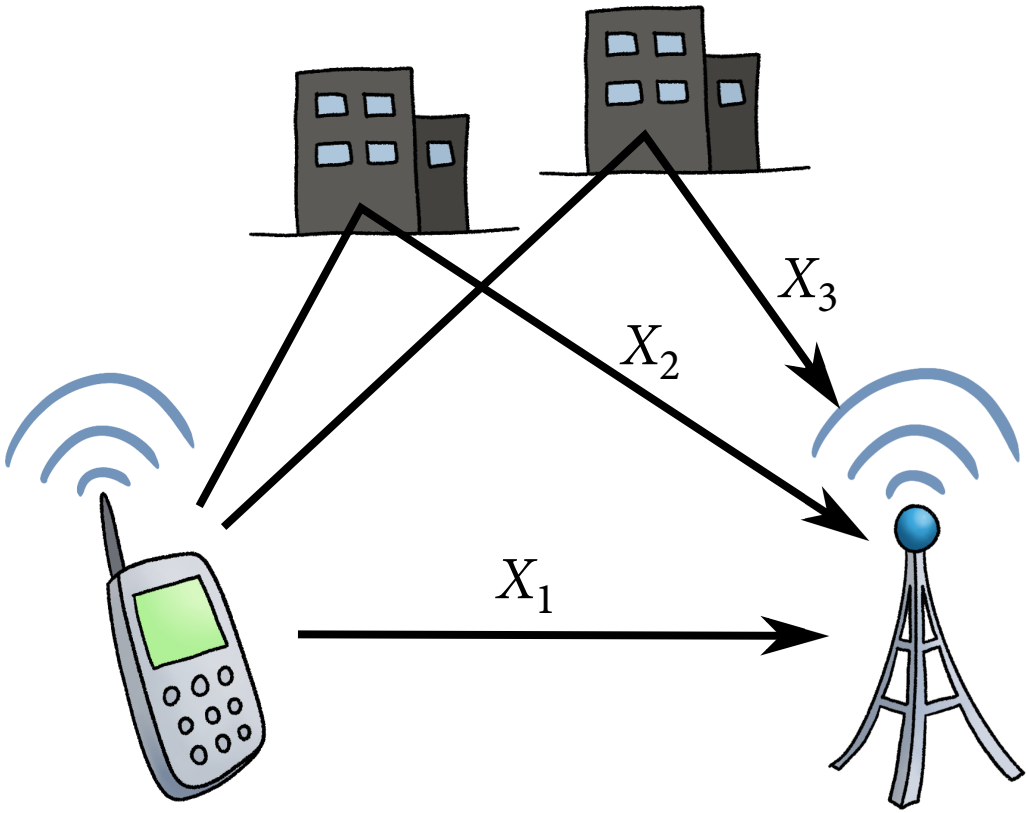
Correlated sources

- Sensor network



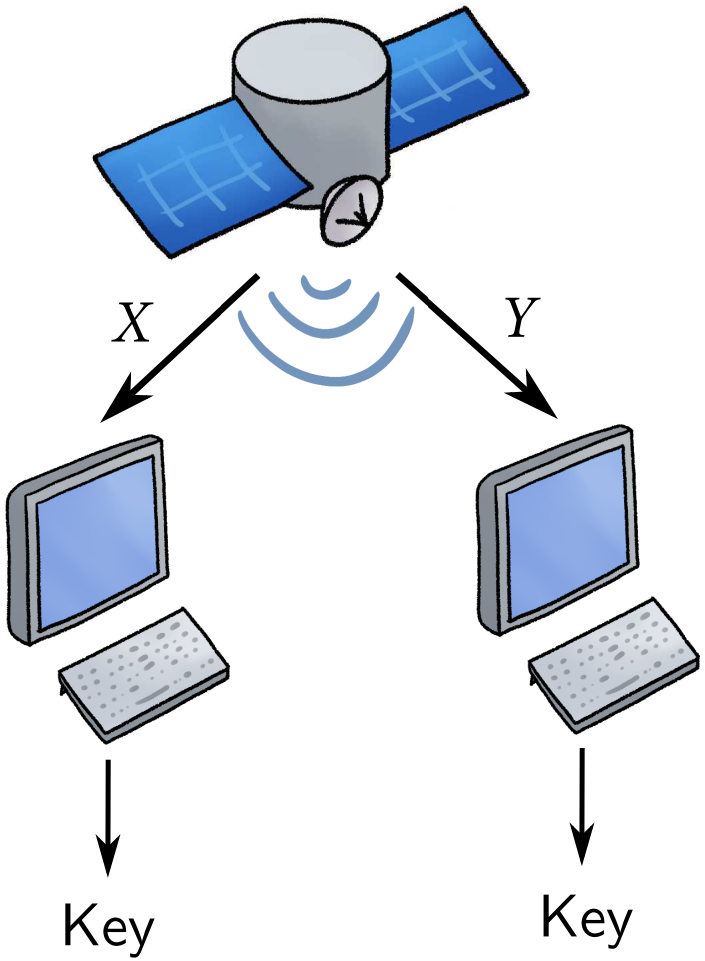
Correlated sources

- Wireless multipath



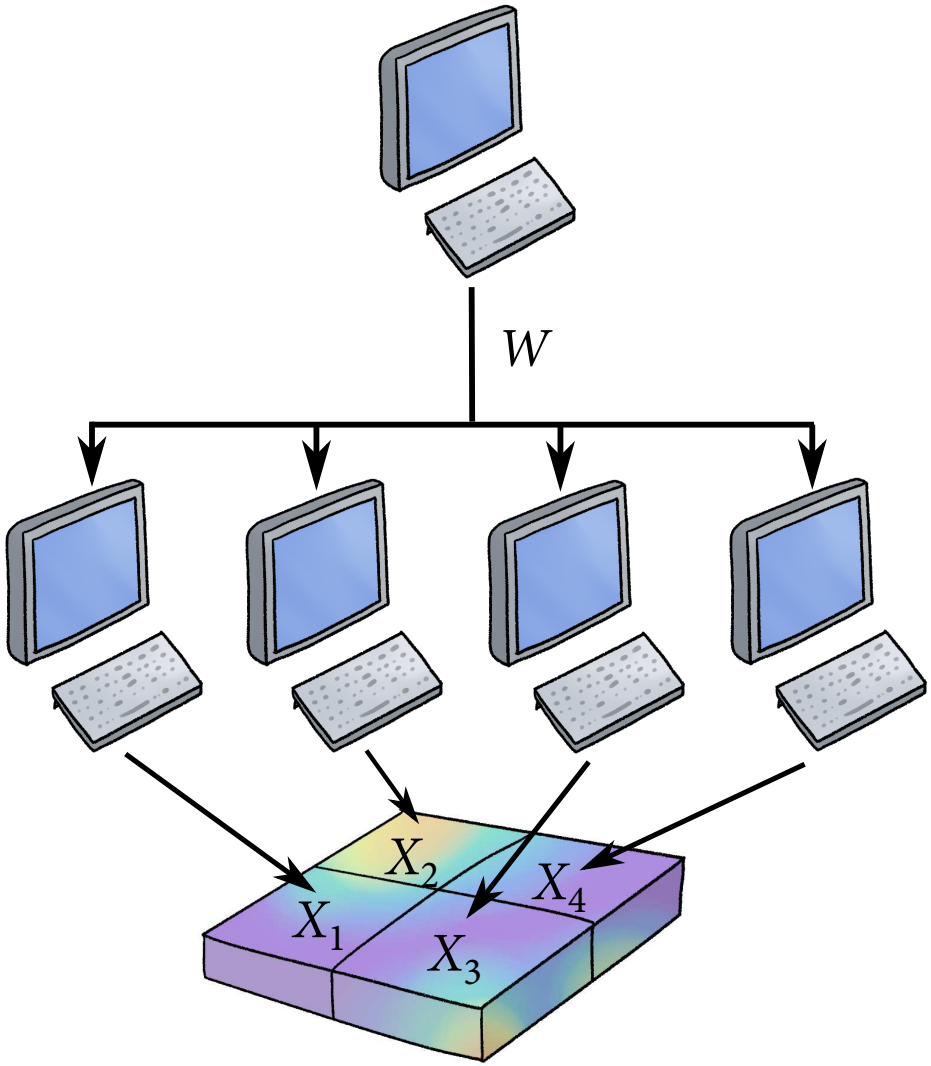
Correlated sources

- Distributed secret key generation



Correlated sources

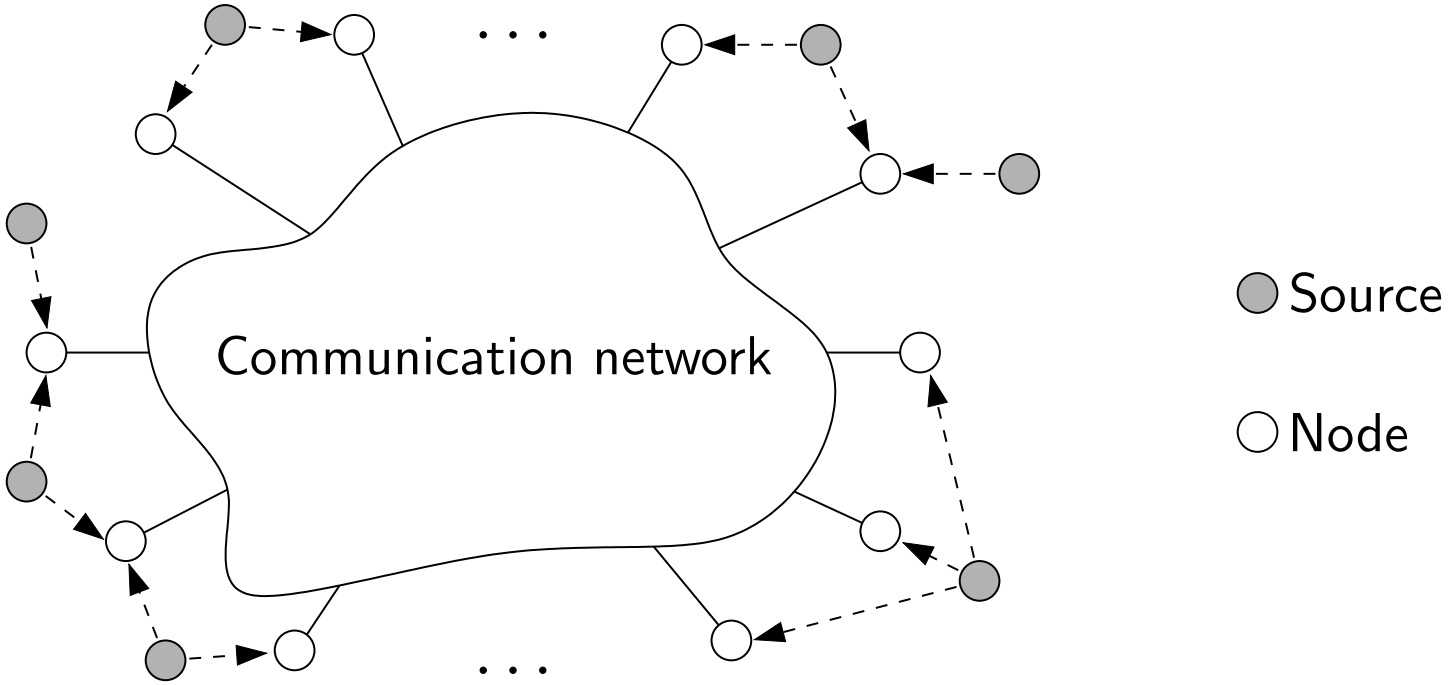
- Distributed simulation



Correlated sources

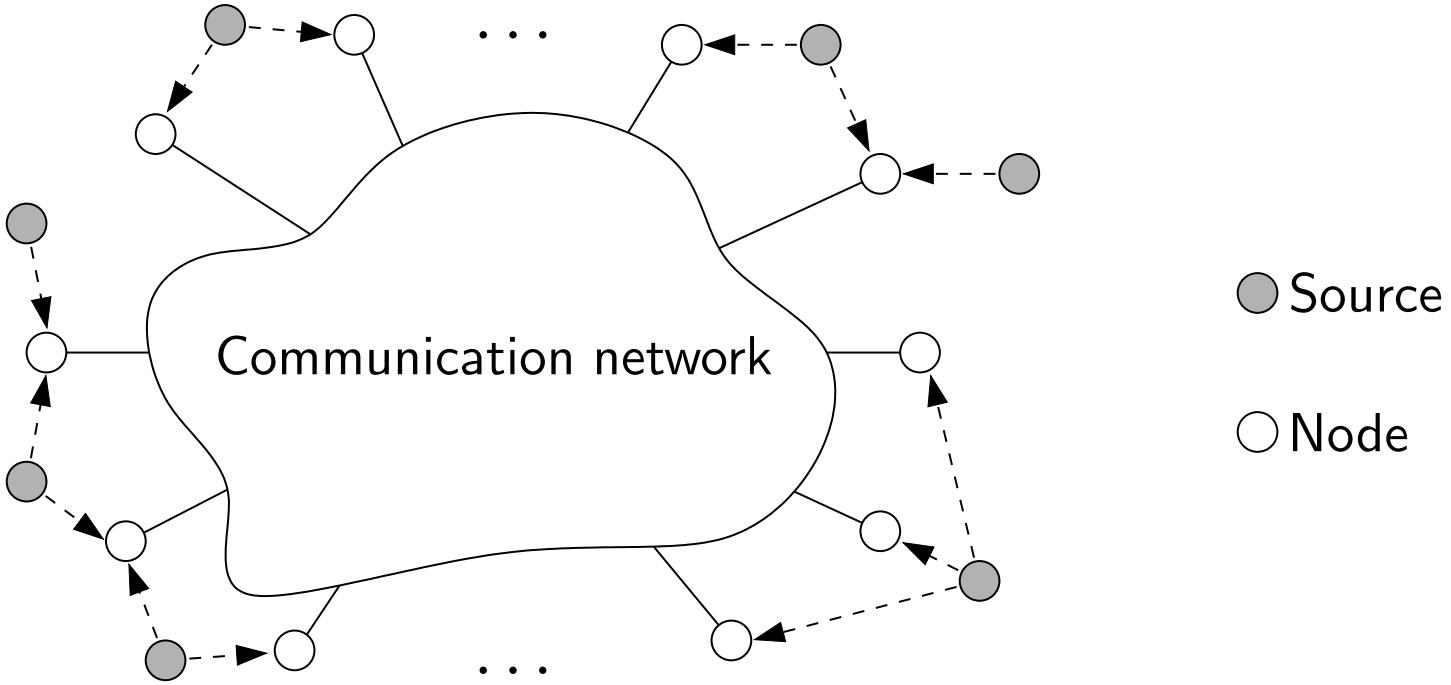
How to measure common information between correlated sources

Network information theory



- Establishes limits on communication/distributed processing in networks

Network information theory



- Establishes limits on communication/distributed processing in networks
- First paper (Shannon 1961): "Two-way communication channels"
- Significant progress in 1970s (Cover 1972, Slepian–Wolf 1973)
- Internet and wireless communication revived interest (EG–Kim 2011)

Measuring common information

- Setups for measuring common information:
  - Distributed compression
  - Distributed key generation
  - Distributed simulation

Measuring common information

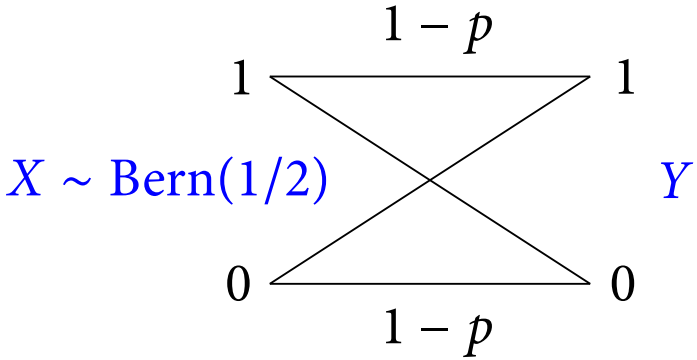
- Setups for measuring common information:
  - Distributed compression
  - Distributed key generation
  - Distributed simulation
- Assume a 2-discrete memoryless source $(\mathcal{X} \times \mathcal{Y}, p(x, y))$ (2-DMS (X, Y))

  (X, Y) generates i.i.d. sequence $(X_1, Y_1), (X_2, Y_2), \ldots \sim p(x, y)$

Measuring common information

- Setups for measuring common information:
  - Distributed compression
  - Distributed key generation
  - Distributed simulation
- Assume a 2-discrete memoryless source $(\mathcal{X} \times \mathcal{Y}, p(x, y))$ (2-DMS (X, Y))

  (X, Y) generates i.i.d. sequence $(X_1, Y_1), (X_2, Y_2), \ldots \sim p(x, y)$
- Examples:



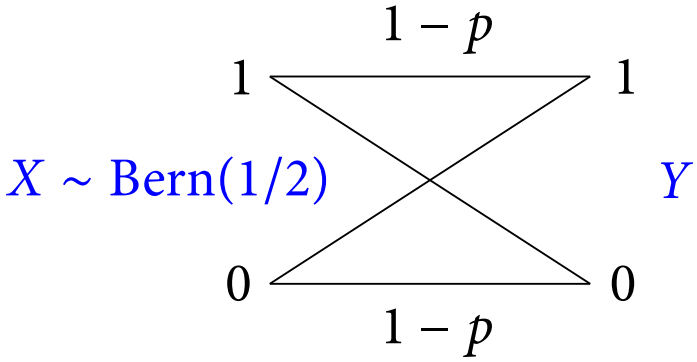
Doubly symmetric binary source

DSBS

Measuring common information

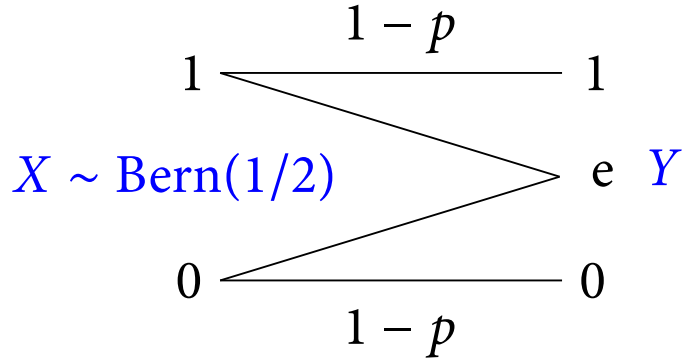
- Setups for measuring common information:
  - Distributed compression
  - Distributed key generation
  - Distributed simulation
- Assume a 2-discrete memoryless source $(\mathcal{X} \times \mathcal{Y}, p(x, y))$ (2-DMS (X, Y))

  (X, Y) generates i.i.d. sequence $(X_1, Y_1), (X_2, Y_2), \ldots \sim p(x, y)$
- Examples:



Doubly symmetric binary source

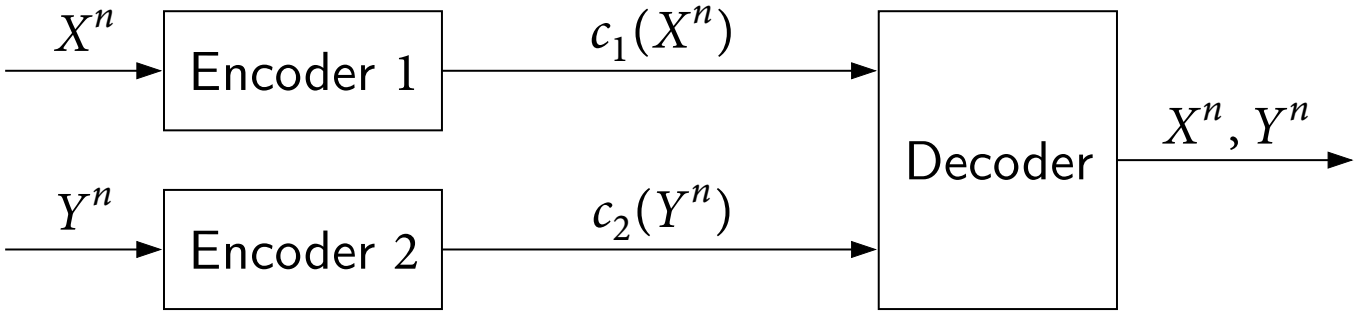
DSBS



Symmetric binary erasure source

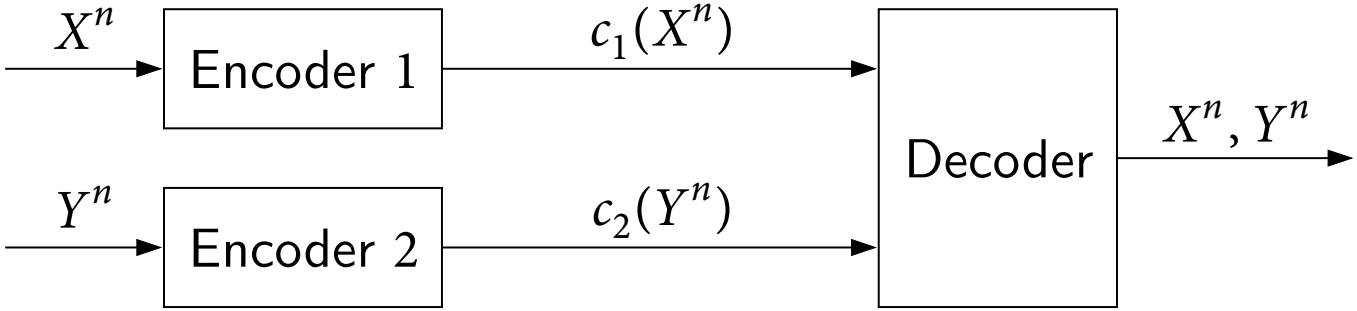
SBES

Distributed zero-error compression



- Let (X, Y) be a 2-DMS; prefix-free codes $c_1(x^n)$, $c_2(y^n)$
- Let L_{jn} be length of c_j and $R_{jn} = (1/n)\,\mathsf{E}(L_{jn})$, $j = 1, 2$

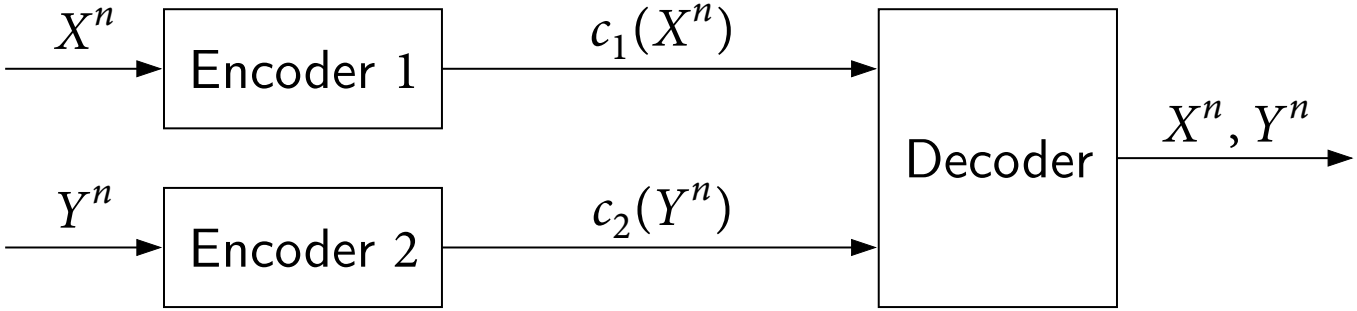
Distributed zero-error compression



- Let (X, Y) be a 2-DMS; prefix-free codes $c_1(x^n)$, $c_2(y^n)$
- Let L_{jn} be length of c_j and $R_{jn} = (1/n)\,\mathsf{E}(L_{jn})$, $j = 1, 2$
- Measure the common information rate between X and Y by

$$R_{\mathrm{C}}^* = H(X) + H(Y) - \inf_n \min_{\text{codes}} (R_{1n} + R_{2n}) \text{ bits/symbol-pair}$$

Distributed zero-error compression



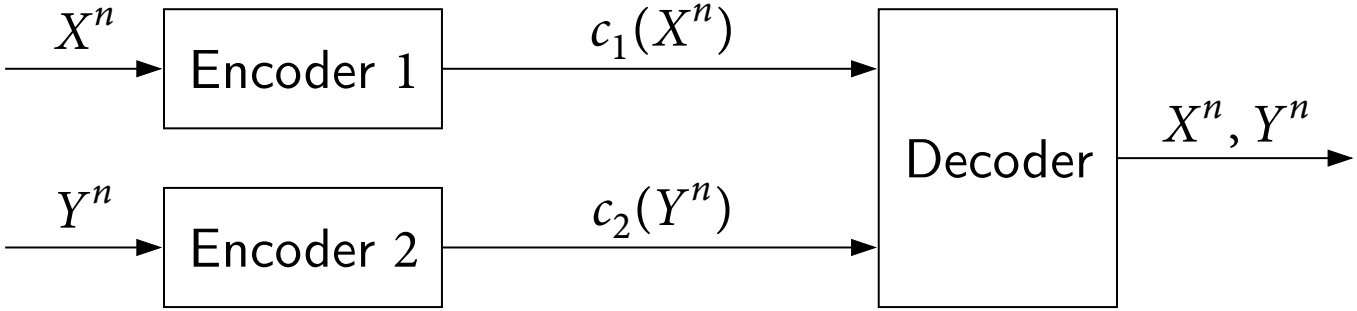
- Let (X, Y) be a 2-DMS; prefix-free codes $c_1(x^n)$, $c_2(y^n)$
- Let L_{jn} be length of c_j and $R_{jn} = (1/n)\,\mathsf{E}(L_{jn})$, $j = 1, 2$
- Measure the common information rate between X and Y by

$$R_{\mathrm{C}}^* = H(X) + H(Y) - \inf_n \min_{\text{codes}} (R_{1n} + R_{2n}) \text{ bits/symbol-pair}$$

- Example: $X = (U, W)$, $Y = (V, W)$, where U, V, W are independent:

$$R_{\mathrm{C}}^* = (H(U) + H(W)) + (H(V) + H(W)) - (H(U) + H(V) + H(W))$$

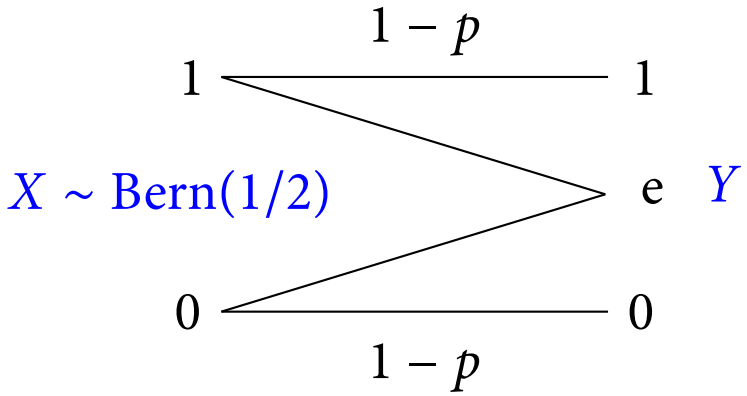
Distributed zero-error compression



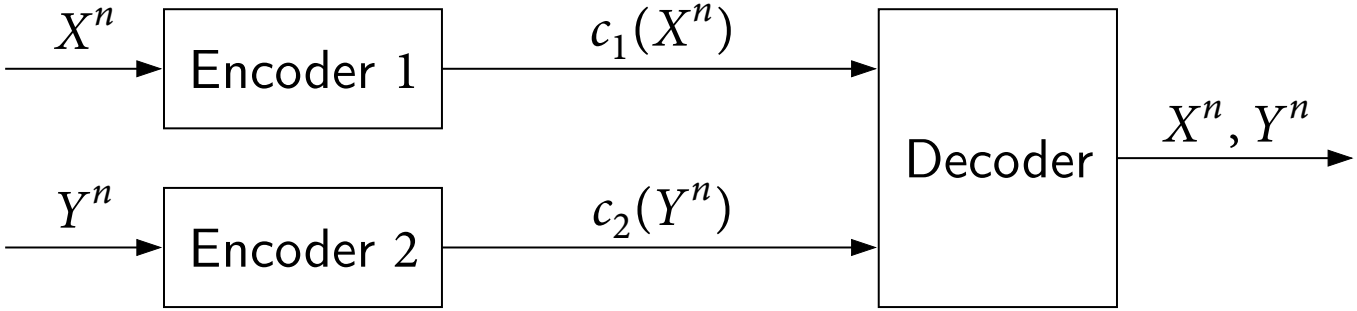
- Measure the common information rate between X and Y by

$$R_C^* = H(X) + H(Y) - \inf_n \min_{\text{codes}} (R_{1n} + R_{2n}) \text{ bits/symbol-pair}$$

- Example: SBES $(H(X) = 1,\ H(Y) = (1-p) + H(p))$



Distributed zero-error compression



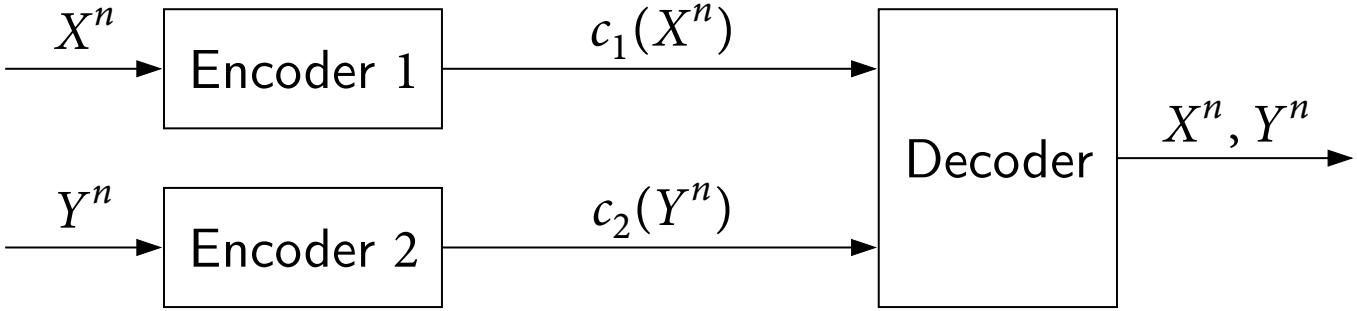
- Example: SBES $\big(H(X) = 1,\ H(Y) = (1-p) + H(p)\big)$

  Can show that $R_{\mathrm{C}}^* = 1 - p$:

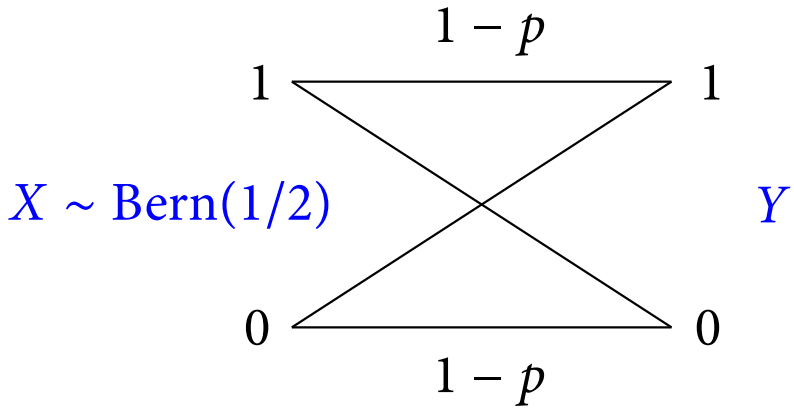
  - Encoder 1 sends X^n: $R_{1n} = 1$ bit/symbol
  - Encoder 2 sends erasure location sequence: $\inf_n R_{2n} = H(p)$ bit/symbol

$\Rightarrow\ R_{\mathrm{C}}^* \le H(X) + H(Y) - (R_{1n} + \inf_n R_{2n}) = 1 - p$

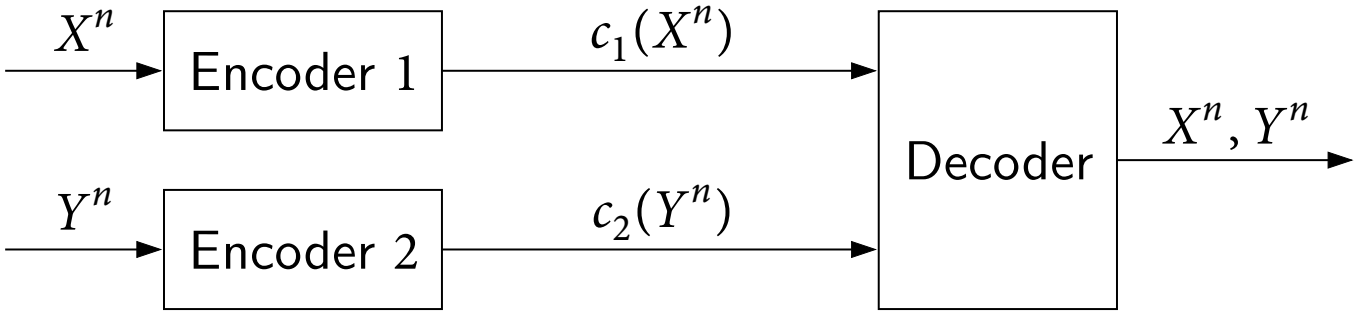
Distributed zero-error compression



- Example: For SBES, $R_{\mathrm{C}}^* = 1 - p$
- Example: For DSBS, $R_{\mathrm{C}}^* = 0$ for $0 < p < 1$



Distributed zero-error compression



- Example: For SBES, $R_{\mathrm{C}}^* = 1 - p$
- Example: For DSBS, $R_{\mathrm{C}}^* = 0$ for $0 < p < 1$
- No computable expression for R_{C}^* is known (Körner–Orlitsky 1998)

Distributed secret key generation

- (X, Y) is a 2-DMS
- Define key rate as $R_n = (1/n)H(K_n)$ bits/symbol-pair

Distributed secret key generation

- (X, Y) is a 2-DMS
- Define key rate as $R_n = (1/n)H(K_n)$ bits/symbol-pair
- The common information rate: $R^* = \sup_n \max_{\text{extractors}} R_n$

Distributed secret key generation

- (X, Y) is a 2-DMS
- Define key rate as $R_n = (1/n)H(K_n)$ bits/symbol-pair
- The common information rate: $R^* = \sup_n \max_{\text{extractors}} R_n$
- Example: $X = (U, W)$, $Y = (V, W)$, where U, V, W are independent

  Then again $R^* = H(W)$

Distributed secret key generation

- Example: Consider (X, Y) with pmf:

		$W=1$		$W=2$	
	X / Y	1	2	3	4
$W=1$	1	0.1	0.2	0	0
	2	0.1	0.1	0	0
	3	0.1	0.1	0	0
$W=2$	4	0	0	0.2	0.1

Distributed secret key generation

- Example: Consider (X, Y) with pmf:

		$W=1$		$W=2$	
	X / Y	1	2	3	4
$W=1$	1	0.1	0.2	0	0
	2	0.1	0.1	0	0
	3	0.1	0.1	0	0
$W=2$	4	0	0	0.2	0.1

- Alice and Bob can agree on i.i.d. key $W^n \Rightarrow R_n = H(W)$

Distributed secret key generation

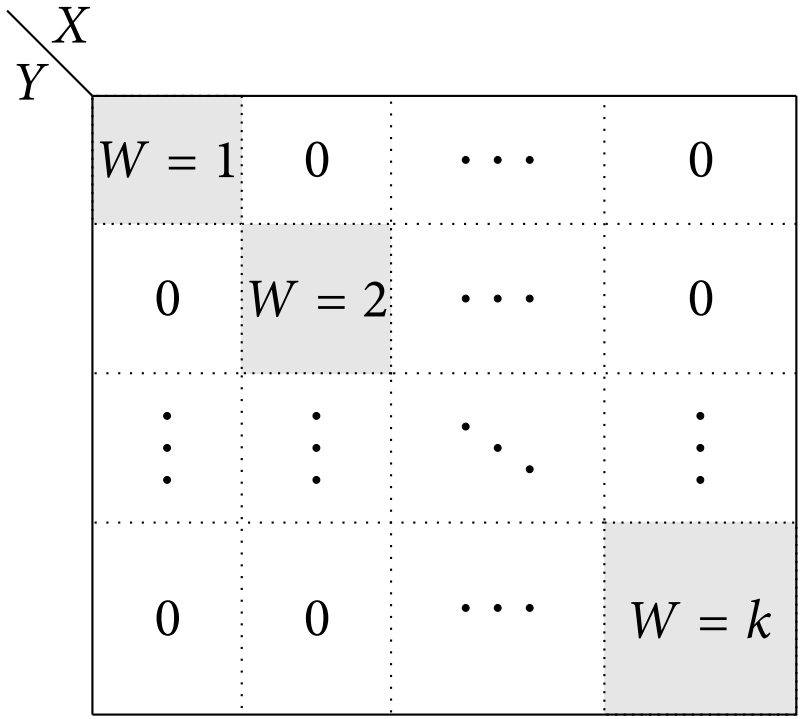
- Example: Consider (X, Y) with pmf:

		$W=1$		$W=2$	
	X / Y	1	2	3	4
$W=1$	1	0.1	0.2	0	0
	2	0.1	0.1	0	0
	3	0.1	0.1	0	0
$W=2$	4	0	0	0.2	0.1

- Alice and Bob can agree on i.i.d. key $W^n \Rightarrow R_n = H(W)$
- Turns out: $R^* = H(W) = 0.881$ bits/symbol-pair

Gács–Körner (1973), Witsenhausen (1975) common information

- Arrange $p(x, y)$ in block diagonal form with largest # of blocks:



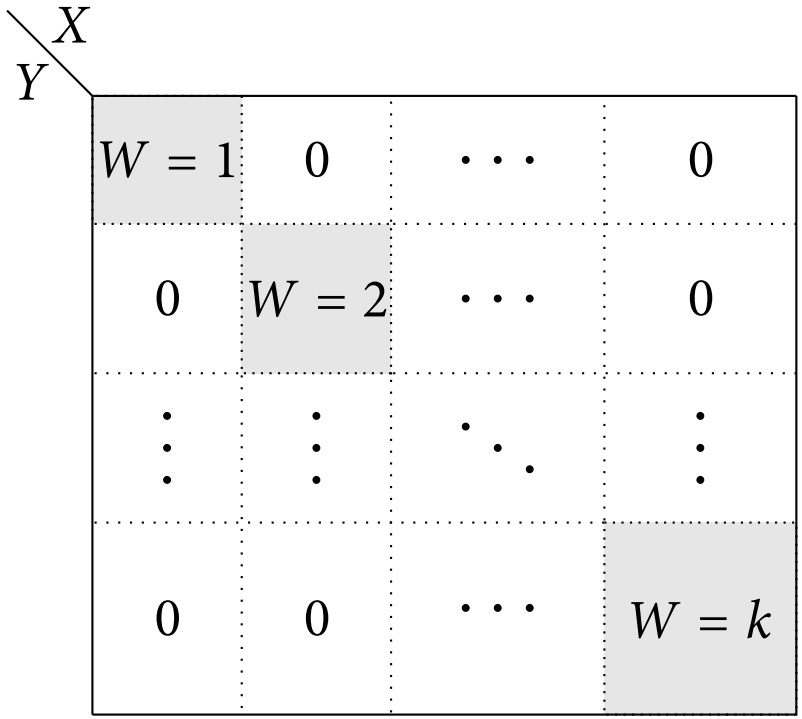
- W is called the common part between X and Y

  It is the highest entropy r.v. that X and Y can agree on

  $K(X; Y) = H(W)$ is called GKW common information

Gács–Körner (1973), Witsenhausen (1975) common information

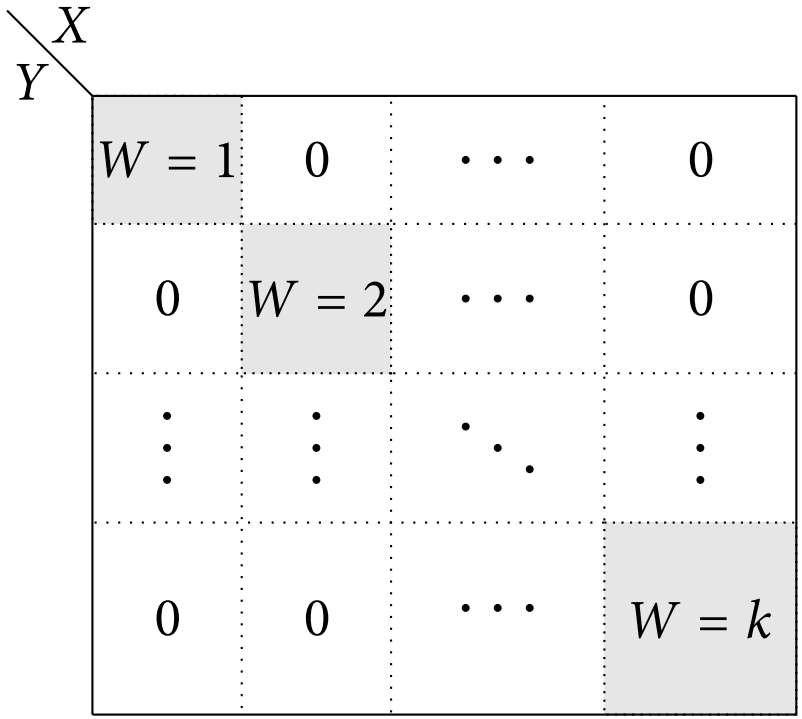
- Arrange $p(x, y)$ in block diagonal form with largest # of blocks:



- W is called the common part between X and Y
- The common part between X^n and Y^n is W^n

Gács–Körner (1973), Witsenhausen (1975) common information

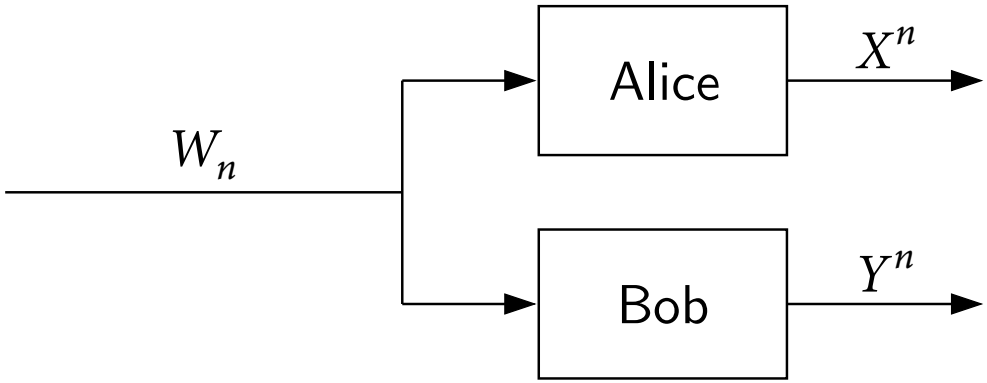
- Arrange $p(x, y)$ in block diagonal form with largest # of blocks:



Common information rate is the GKW common information

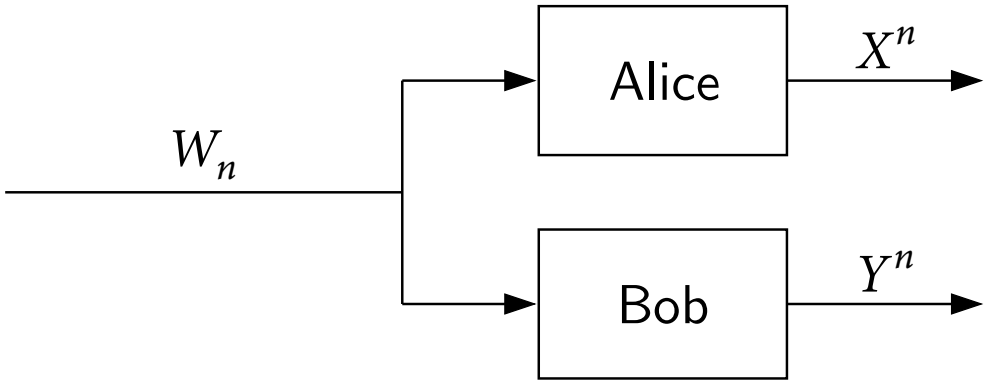
$$R^* = K(X;Y) = H(W)$$

Distributed simulation



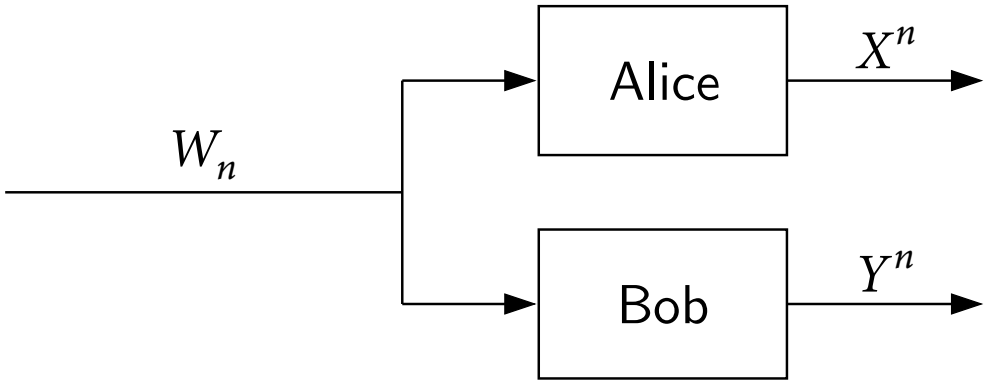
- W_n is a common information random variable
- Wish to generate (X^n, Y^n) i.i.d. $\sim p(x, y)$

Distributed simulation



- W_n is a common information random variable
- Wish to generate (X^n, Y^n) i.i.d. $\sim p(x, y)$
- Define the distributed simulation rate $R_n = (1/n)H(W_n)$
- The common information rate: $R_S^* = \inf_n \min_{W_n, \text{generators}} R_n$

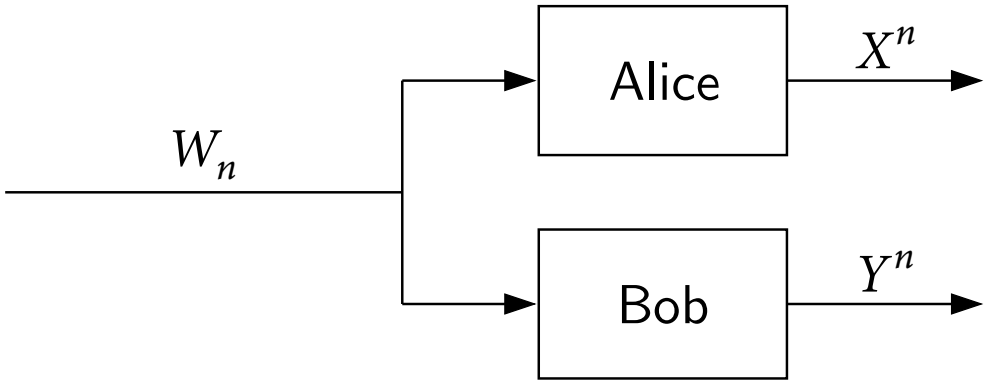
Distributed simulation



- W_n is a common information random variable
- Wish to generate (X^n, Y^n) i.i.d. $\sim p(x, y)$
- Define the distributed simulation rate $R_n = (1/n)H(W_n)$
- The common information rate: $R_S^* = \inf_n \min_{W_n, \text{generators}} R_n$
- Example: $X = (U, W)$, $Y = (V, W)$, where U, V, W are independent

  Then again $R_S^* = H(W)$ (set $W_n = W^n$)

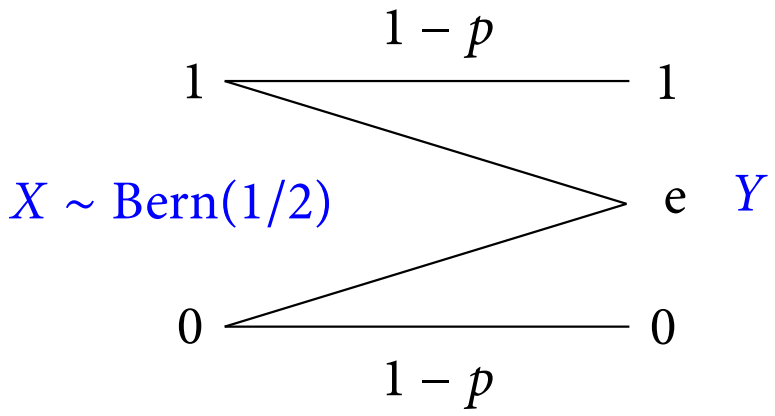
Distributed simulation



- W_n is a common information random variable
- Wish to generate (X^n, Y^n) i.i.d. $\sim p(x, y)$
- Define the distributed simulation rate $R_n = (1/n)H(W_n)$
- The common information rate: $R_S^* = \inf_n \min_{W_n, \text{generators}} R_n$
- No computable expression for R_S^* is known

Distributed simulation

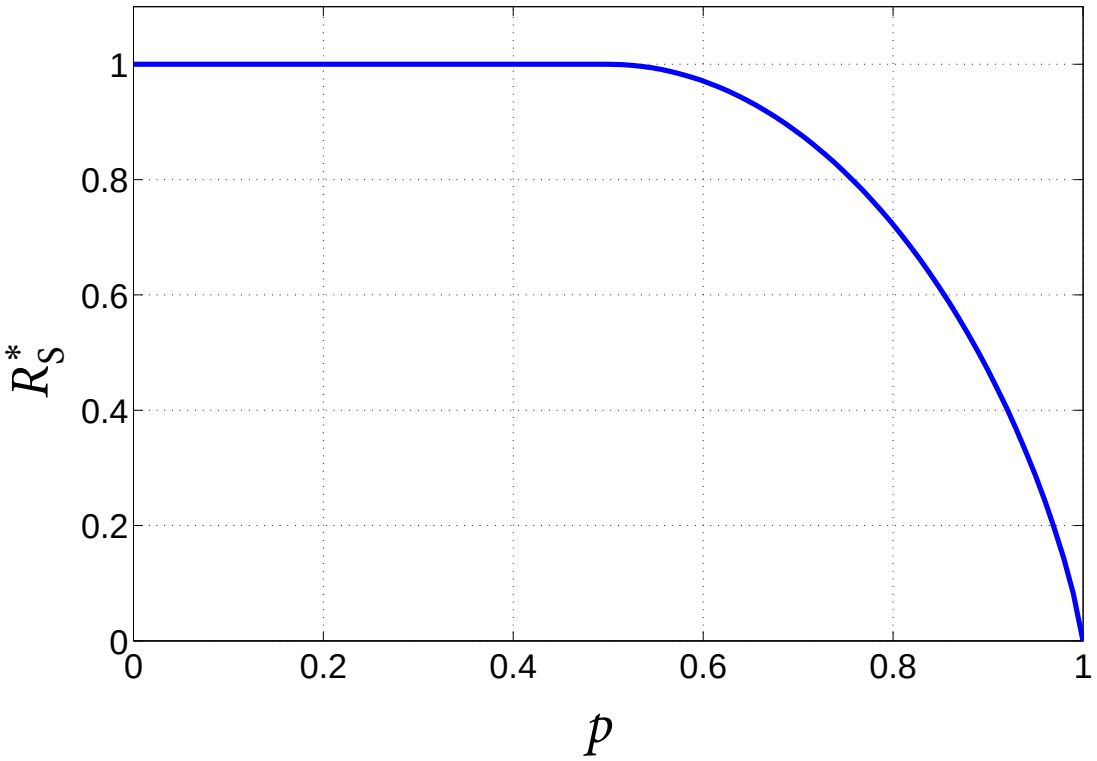
- Example: SBES (Kumar–Li–EG, ISIT 2014)



Distributed simulation

- Example: SBES (Kumar–Li–EG, ISIT 2014)

$$R_{\mathrm{S}}^* = \begin{cases} 1 & \text{if } p \le 1/2 \\ H(p) & \text{if } p > 1/2 \end{cases}$$



Summary

- Entropy is a "universal" measure of information

Summary

- Entropy is a "universal" measure of information
- For $X = (U, W)$, $Y = (V, W)$, where U, V, W are independent
  - Common information is measured by $H(W)$

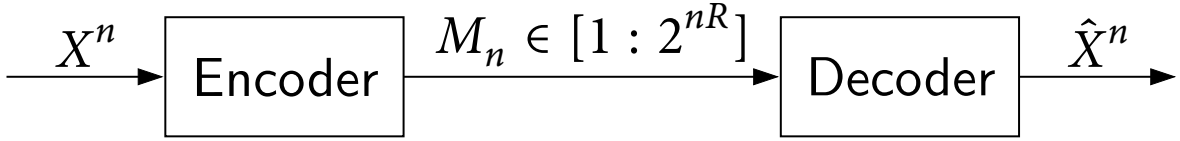
Summary

- Entropy is a "universal" measure of information
- In general, common information has several measures:
  - For zero-error distributed compression: R_{C}^* (no computable expression)
  - For distributed key generation: $K(X;Y)$ (GKW common information)
  - For distributed simulation: R_{S}^* (no computable expression)

Summary

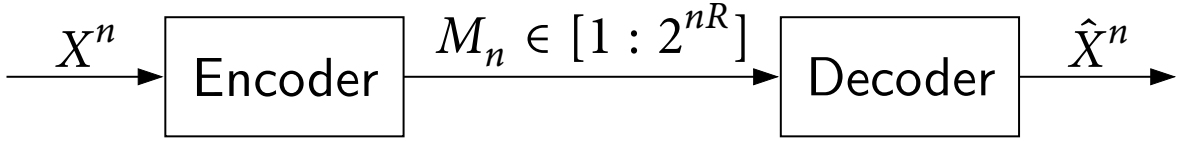
- Entropy is a "universal" measure of information
- In general, common information has several measures:
  - For zero-error distributed compression: R_{C}^{*} (no computable expression)
  - For distributed key generation: $K(X;Y)$ (GKW common information)
  - For distributed simulation: R_{S}^{*} (no computable expression)
- We can say more by considering relaxed versions of these setups
  - Distributed lossless compression
  - Distributed approximate simulation
- Common information for approximate key generation same as for exact

Lossless compression



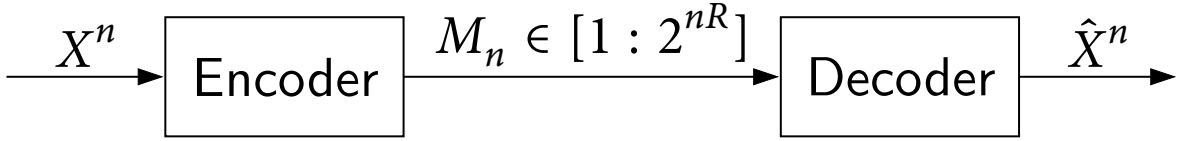
- Let X be a DMS
- Use fixed-length (block) codes ($M_n \in \{1, \ldots, 2^{nR}\}$ is an nR-bit sequence)
- Probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{X}^n \neq X^n\}$

Lossless compression



- Let X be a DMS
- Use fixed-length (block) codes ($M_n \in \{1, \ldots, 2^{nR}\}$ is an nR-bit sequence)
- Probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{X}^n \neq X^n\}$
- R is achievable if $\exists$ a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$
- The information rate of X: $R^* = \inf\{R : R \text{ is achievable}\}$

Lossless compression



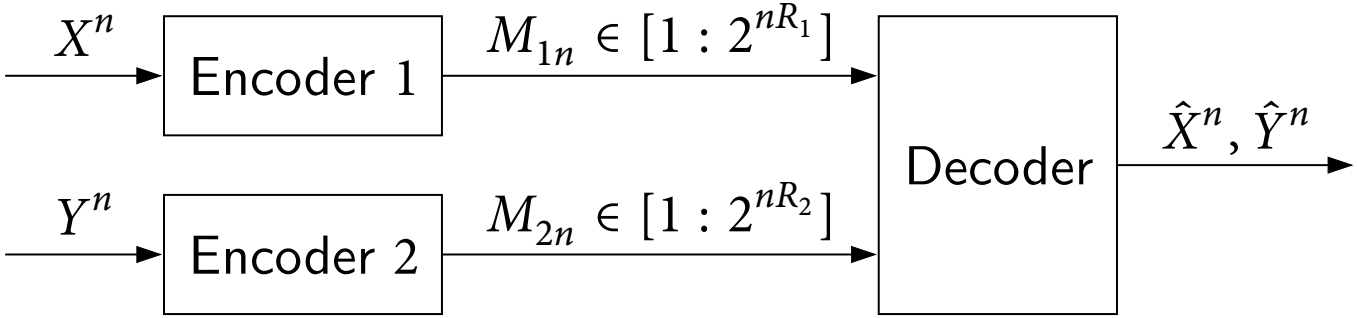
- Let X be a DMS
- Use fixed-length (block) codes ($M_n \in \{1, \ldots, 2^{nR}\}$ is an nR-bit sequence)
- Probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{X}^n \neq X^n\}$
- R is achievable if $\exists$ a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$
- The information rate of X: $R^* = \inf\{R : R \text{ is achievable}\}$

Information rate is the entropy (Shannon 1948)

$$R^* = H(X)$$

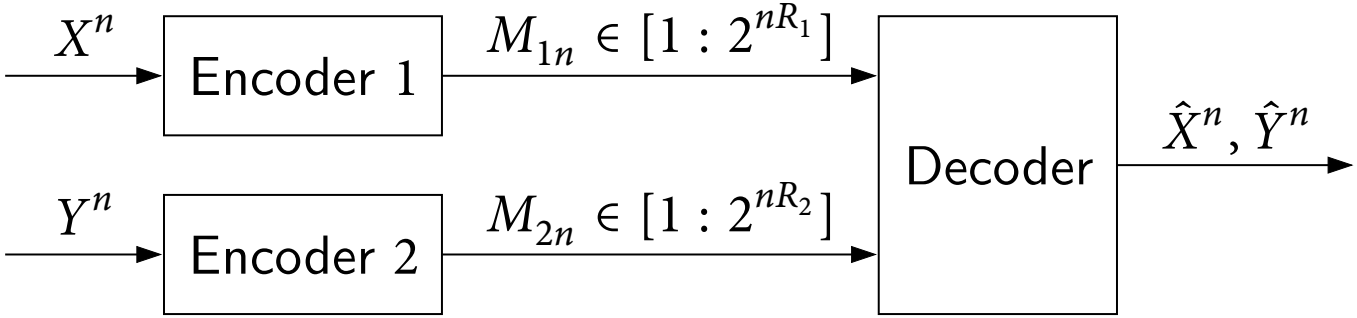
- Same as in zero-error!

Distributed lossless compression



- Again assume fixed length codes
- Probability of error: $P_e^{(n)} = \mathrm{P}\{(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\}$
- (R_1, R_2) achievable if $\exists$ a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$

Distributed lossless compression



- Again assume fixed length codes
- Probability of error: $P_e^{(n)} = \mathrm{P}\{(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\}$
- (R_1, R_2) achievable if $\exists$ a sequence of codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$
- The common information rate

$$R^* = H(X) + H(Y) - \inf\{R_1 + R_2 : (R_1, R_2) \text{ is achievable}\}$$

Distributed lossless compression

- The common information rate

$$R^* = H(X) + H(Y) - \inf\{R_1 + R_2 : (R_1, R_2) \text{ is achievable}\}$$

Common information rate is mutual information (Slepian–Wolf 1973)

$$R^* = I(X;Y) = H(X) + H(Y) - H(X,Y)$$

- The same compression rate as if X and Y are compressed together!

Distributed lossless compression

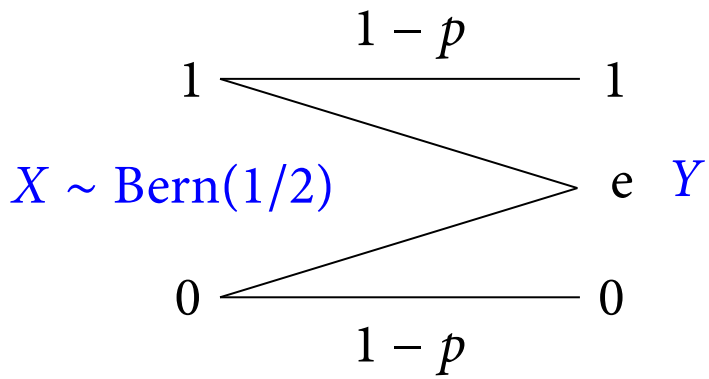
- The common information rate

$$R^* = H(X) + H(Y) - \inf\{R_1 + R_2 : (R_1, R_2) \text{ is achievable}\}$$

Common information rate is mutual information (Slepian–Wolf 1973)

$$R^* = I(X;Y) = H(X) + H(Y) - H(X,Y)$$

- Example: (X, Y) is SBES, $I(X;Y) = 1 - p = R_C^*$ (zero-error)



Distributed lossless compression

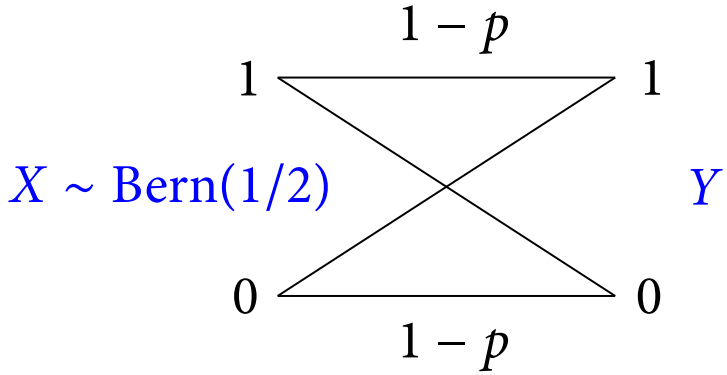
- The common information rate

$$R^* = H(X) + H(Y) - \inf\{R_1 + R_2 : (R_1, R_2) \text{ is achievable}\}$$

Common information rate is mutual information (Slepian–Wolf 1973)

$$R^* = I(X;Y) = H(X) + H(Y) - H(X,Y)$$

- Example: (X, Y) is SBES, $I(X;Y) = 1 - p = R_C^*$ (zero-error)
- Example: (X, Y) is DSBS, $I(X;Y) = 1 - H(p)$, $R_C^* = 0$, $0 < p < 1$



Distributed lossless compression

- The common information rate

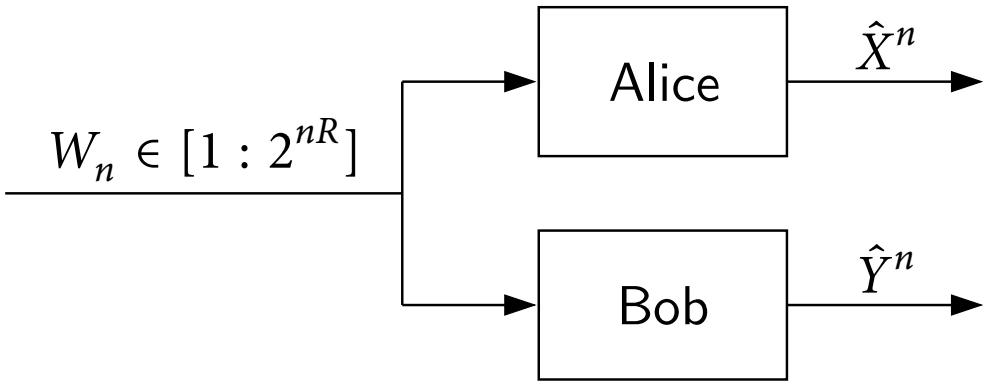
$$R^* = H(X) + H(Y) - \inf\{R_1 + R_2 : (R_1, R_2) \text{ is achievable}\}$$

Common information rate is mutual information (Slepian–Wolf 1973)

$$R^* = I(X;Y) = H(X) + H(Y) - H(X,Y)$$

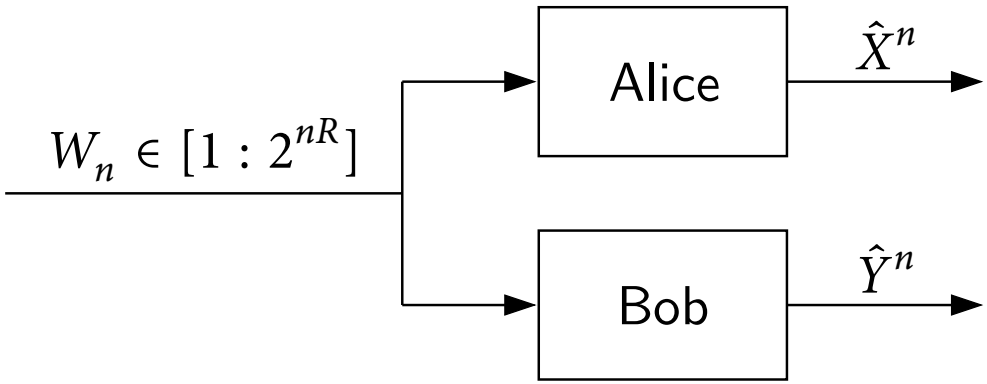
- Example: (X, Y) is SBES, $I(X;Y) = 1 - p = R_C^*$ (zero-error)
- Example: (X, Y) is DSBS, $I(X;Y) = 1 - H(p)$, $R_C^* = 0$, $0 < p < 1$
- Also, $I(X;Y) = 0$ only if X and Y are independent
- Hence, captures dependence better than previous measures

Distributed approximate simulation



- W_n randomly distributed over $[1 : 2^{nR}]$ (random nR-bit sequence)
- Wish to simulate (X^n, Y^n) i.i.d. $\sim p(x, y)$ approximately

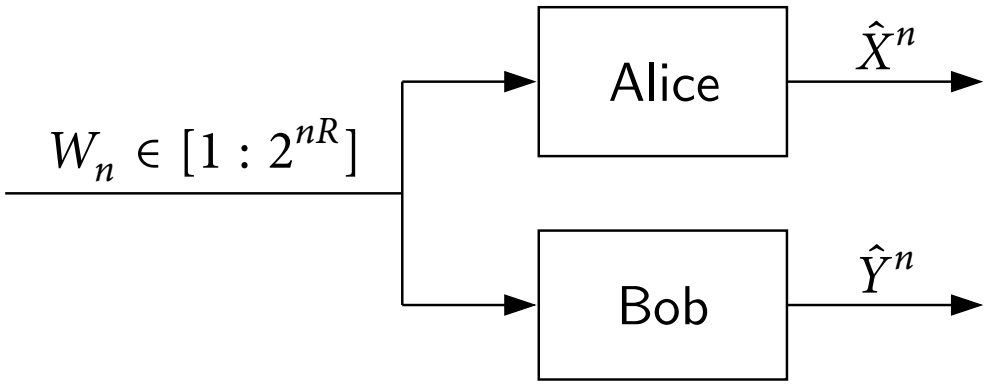
Distributed approximate simulation



- W_n randomly distributed over $[1 : 2^{nR}]$ (random nR-bit sequence)
- Wish to simulate (X^n, Y^n) i.i.d. $\sim p(x, y)$ approximately
- Total variation distance between $(\hat{X}^n, \hat{Y}^n)$ and (X^n, Y^n):

$$d_n = \sum_{(x^n, y^n)} \left| p_{\hat{X}^n, \hat{Y}^n}(x^n, y^n) - \prod_{i=1}^{n} p_{X,Y}(x_i, y_i) \right|$$

Distributed approximate simulation



- W_n randomly distributed over $[1 : 2^{nR}]$ (random nR-bit sequence)
- Wish to simulate (X^n, Y^n) i.i.d. $\sim p(x, y)$ approximately
- Total variation distance between $(\hat{X}^n, \hat{Y}^n)$ and (X^n, Y^n):

$$d_n = \sum_{(x^n, y^n)} \left| p_{\hat{X}^n, \hat{Y}^n}(x^n, y^n) - \prod_{i=1}^{n} p_{X,Y}(x_i, y_i) \right|$$

- R is achievable if $\exists$ a sequence of generators with $\lim_{n\to\infty} d_n = 0$
- The common information rate: $R^* = \inf\{R : R \text{ is achievable}\}$

Distributed approximate simulation

Common information rate is Wyner (1975) common information

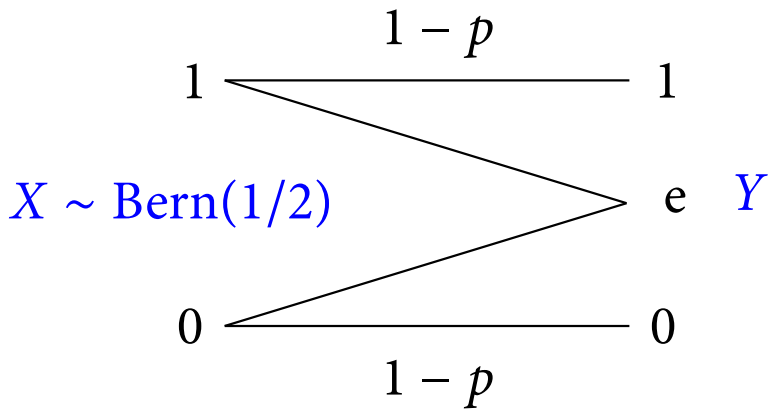
$$R^* = J(X;Y) = \min_{X \to W \to Y} I(X,Y;W)$$

Distributed approximate simulation

Common information rate is Wyner (1975) common information

$$R^* = J(X;Y) = \min_{X \to W \to Y} I(X,Y;W)$$

- Example: SBES

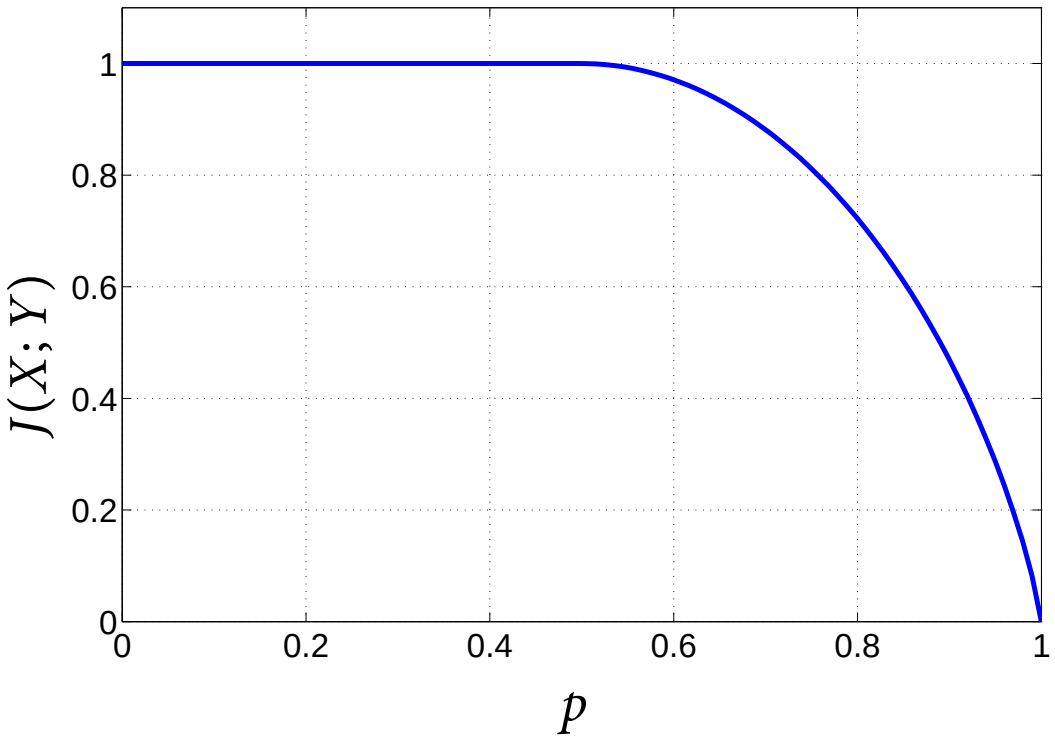


Distributed approximate simulation

Common information rate is Wyner (1975) common information

$$R^* = J(X;Y) = \min_{X \to W \to Y} I(X,Y;W)$$

- Example: SBES

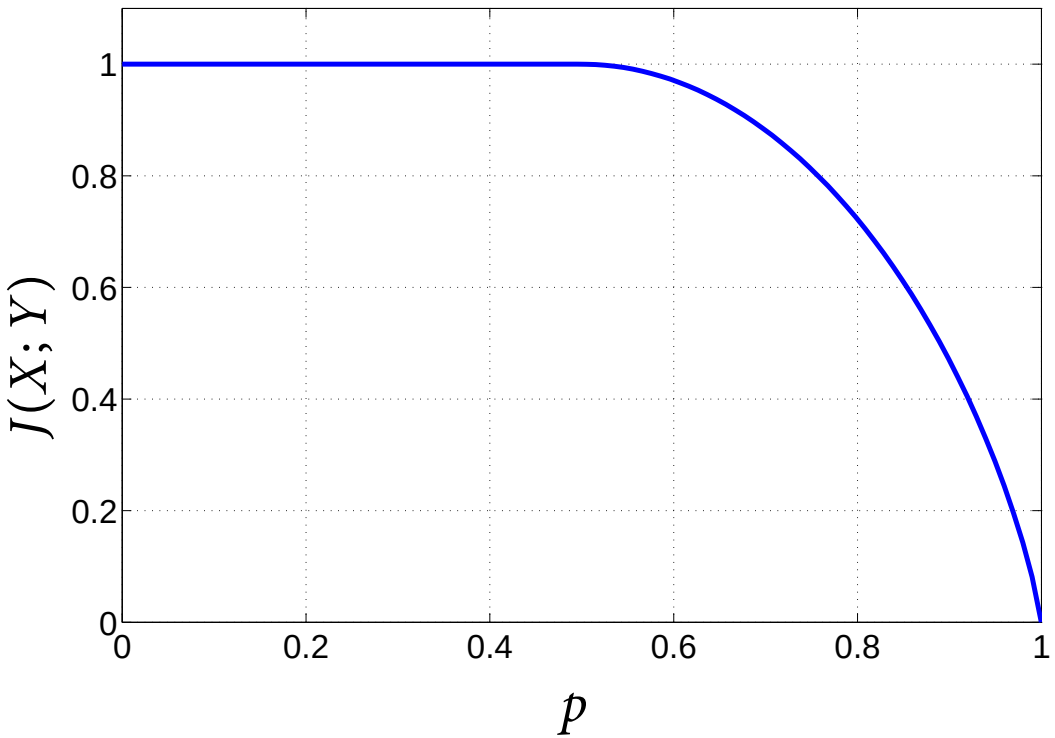


Distributed approximate simulation

Common information rate is Wyner (1975) common information

$$R^* = J(X;Y) = \min_{X \to W \to Y} I(X,Y;W)$$

- Example: SBES, $J(X;Y) = R_S^*$ (exact common information)



Summary

- There are several well-motivated measures of common information

Summary

- There are several well-motivated measures of common information
- For distributed compression:
  - Zero-error: R_{C}^{*} (no computable expression)
  - Lossless: $I(X;Y)$ (mutual information)

Summary

- There are several well-motivated measures of common information
- For distributed compression:
  - Zero-error: R_{C}^{*} (no computable expression)
  - Lossless: $I(X;Y)$ (mutual information)
- For distributed key generation: $K(X;Y)$ (GKW common information)
  - Same for approximate key agreement

Summary

- There are several well-motivated measures of common information
- For distributed compression:
  - Zero-error: R_{C}^* (no computable expression)
  - Lossless: $I(X;Y)$ (mutual information)
- For distributed key generation: $K(X;Y)$ (GKW common information)
  - Same for approximate key agreement
- For distributed simulation:
  - Exact simulation: R_{S}^* (no computable expression)
  - Approximate simulation: $J(X;Y)$ (Wyner common information)

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

- Common information between X and $Y \le$ information of each
- Strict inequality for SBES:
  - $H(X) = 1$, $H(Y) = (1-p) + H(p)$
  - $R_{\mathrm{S}}^* = H(p)$ if $p > 1/2$

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

- Common information for approximate distributed simulation $\le$ for exact
- Equal for SBES (ISIT 2014)
- Open problem: Are they equal in general?

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

- Strict inequality for SBES:
  - $I(X;Y) = 1 - p$
  - $J(X;Y) = 1$ if $p \le 1/2$

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

- Common information for distributed zero-error compression $\le$ for lossless
- Strict inequality for DSBS:
  - $R_{\mathrm{C}}^* = 0$
  - $I(X;Y) = 1 - H(p)$

Comparing common information measures

$$0 \le K(X;Y) \le R_{\mathrm{C}}^* \le I(X;Y) \le J(X;Y) \le R_{\mathrm{S}}^* \le \min\{H(X), H(Y)\}$$

- Common part needs to be sent only once in distributed compression
- Strict inequality for SBEC:
  - $K(X;Y) = 0$
  - $R_{\mathrm{C}}^* = 1 - p$

Conclusion

- Entropy is a universal measure of information
  - For both zero-error and lossless compression
  - For randomness generation and source simulation

Conclusion

- Entropy is a universal measure of information
  - For both zero-error and lossless compression
  - For randomness generation and source simulation
- The story for common information is much richer:
  - Five well-motivated common information measures

$$K \leq R_{\mathrm{C}}^{*} \leq I \leq J \leq R_{\mathrm{S}}^{*}$$

  - They are not always equal!
  - Some don't seem to be computable in general
  - Different for zero-error distributed compression from lossless
  - Are they also different for exact distributed simulation from approximate?

Acknowledgments

- Work on exact distributed simulation is joint with:
  - Gowtham Kumar
  - Cheuk Ting Li
- Young-Han Kim provided many valuable suggestions

Thank You!

References

Arıkan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory*, 55(7), 3051–3073.

Berrou, C. and Glavieux, A. (1996). Near optimum error correcting coding and decoding: Turbo-codes. *IEEE Trans. Comm.*, 44(10), 1261–1271.

Cover, T. M. (1972). Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1), 2–14.

El Gamal, A. and Kim, Y.-H. (2011). *Network Information Theory*. Cambridge, Cambridge.

Elias, P. (1972). The efficient construction of an unbiased random sequence. *The Annals of Mathematical Statistics*, 43(3), 865–870.

Gács, P. and Körner, J. (1973). Common information is far less than mutual information. *Probl. Control Inf. Theory*, 2(2), 149–162.

Gallager, R. G. (1963). *Low-Density Parity-Check Codes*. MIT Press, Cambridge, MA.

Knuth, D. E. and Yao, A. C. (1976). The complexity of random number generation. In *Algorithms and Complexity (Proc. Symp., Carnegie Mellon Univ., Pittsburgh, Pa., 1976)*, pp. 357–428. Academic Press, New York.

References (cont.)

Körner, J. and Orlitsky, A. (1998). Zero-error information theory. *IEEE Trans. Inf. Theory*, 44(6), 2207–2229.

Orlitsky, A. and Roche, J. R. (2001). Coding for computing. *IEEE Trans. Inf. Theory*, 47(3), 903–917.

Richardson, T. and Urbanke, R. (2008). *Modern Coding Theory*. Cambridge University Press, Cambridge.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3), 379–423, 27(4), 623–656.

Shannon, C. E. (1961). Two-way communication channels. In *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. I, pp. 611–644. University of California Press, Berkeley.

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

Ungerboeck, G. (1982). Channel coding with multilevel/phase signals. *IEEE Trans. Info. Theory*, 28(1), 55–67.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory*, 13(2), 260–269.

References (cont.)

Witsenhausen, H. S. (1975). On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.*, 28(1), 100–113.

Wyner, A. D. (1975). The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2), 163–179.

Ziv, J. and Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory*, 23(3), 337–343.

Ziv, J. and Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory*, 24(5), 530–536.