CS3236 Introduction to Information theory

Lecture 3: Source Coding Theorem

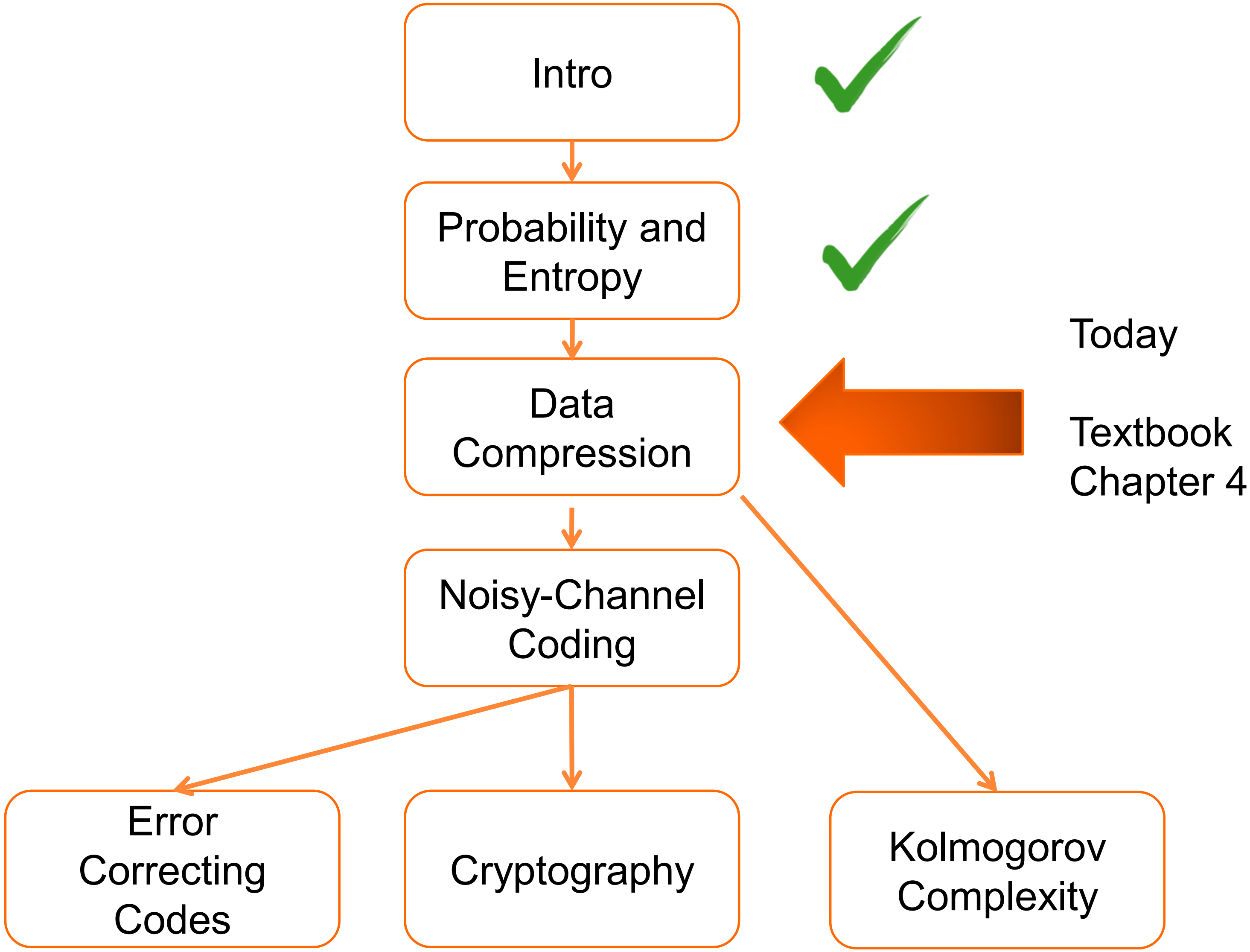
Course given by Pierre Senellart
Material by Stephanie Wehner, with additions by P. Senellart

LET'S RECAP…

Class survival guide

- Lecture on Monday 2pm-3:30pm
- Tutorial on Wed 4-5pm and Thu 2-3pm, starting this week, COM1-02-18
  - Choose one session, no need to go to both
  - Discussion, QA, help with homework exercises, projects
- Recess week September 20-28
- No physical lecture/tutorial on Sep 8-12, e-Learning material
- No lecture on October 6 (Hari Raya Haji)
- No tutorial on week 10, October 22-23 (Deepavali)
- Available to meet for answering individual questions. Use http://www.doodle.com/psenellart to schedule a meeting. Office Icube #03-09.
- Grading
  - 50% Exam (Friday November 28)
  - 50% Continuous Assessment
    - 10% Mid-Term exam (October 13)
    - 20% Homework (assigned each week, due next Monday)
    - 20% Small project (handed out September 1, due November 14)
- If you have any doubt about the schedule, check IVLE

WHERE DO WE GO FROM HERE?



WHAT WE DID LAST TIME

- Probability 101 (see chapter 2)

- Weak law of large numbers (chapter 4.5)

  IID random variables $X^N = X_1, \ldots,$

$$P((Y - \mathbb{E}(Y))^2 \geq \alpha) \leq \frac{\mathrm{Var(X_j)}}{\alpha N} \qquad Y = \frac{1}{N}\sum_{j=1}^{N} X_j$$

- Entropy: Average information content

$$H(X) = \sum_x P(x)h(x) = \sum_x P(x) \log \frac{1}{P(x)} = -\sum_x P(x) \log P(x)$$

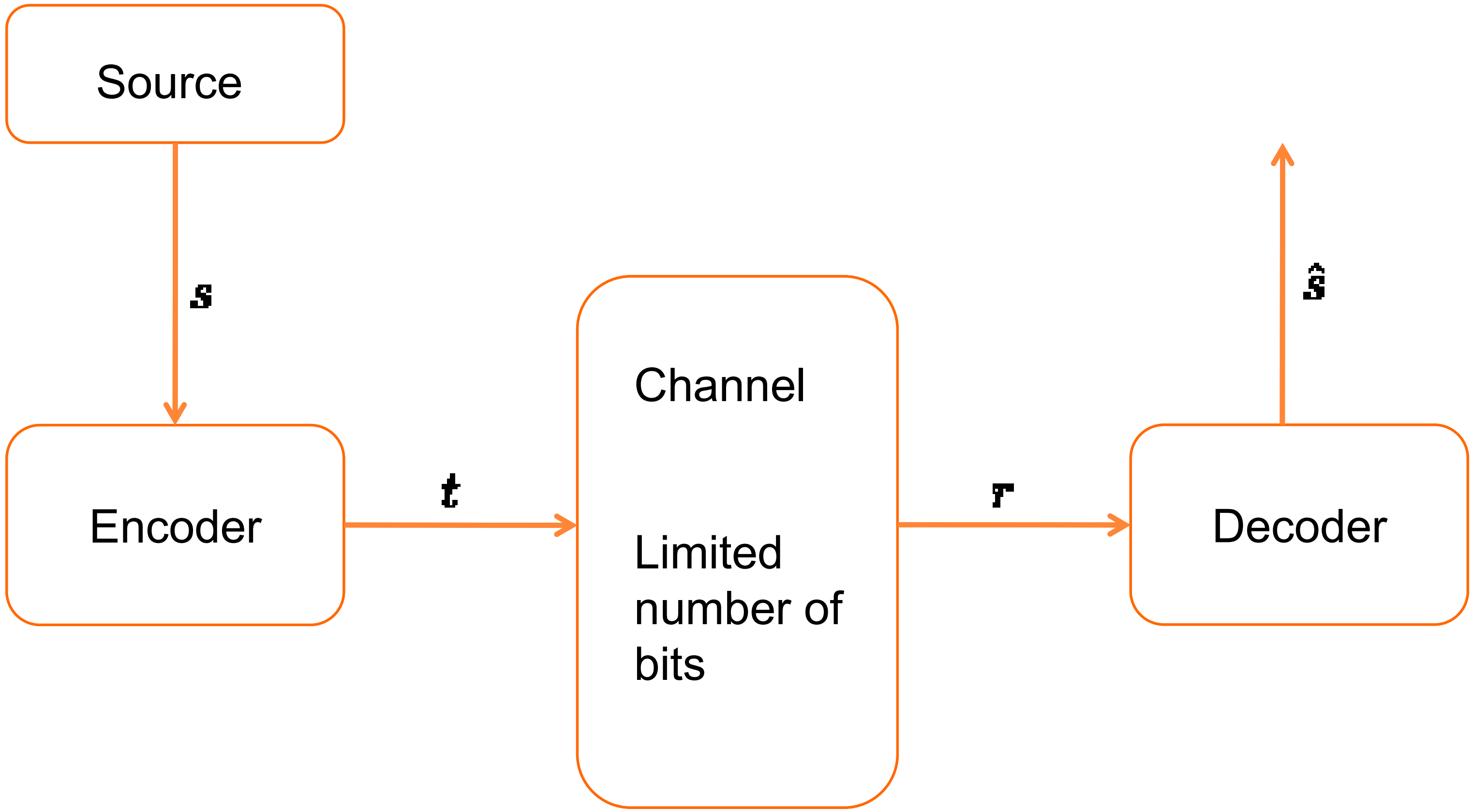
- X,Y independent variables

$$H(X,Y) = H(X) + H(Y)$$

WHAT WE'LL DO TODAY

- Data compression
- Lossless vs lossy compression
- Shannon's source coding theorem: how well can we compress information (in a lossy scenario)?
- Proof of this theorem

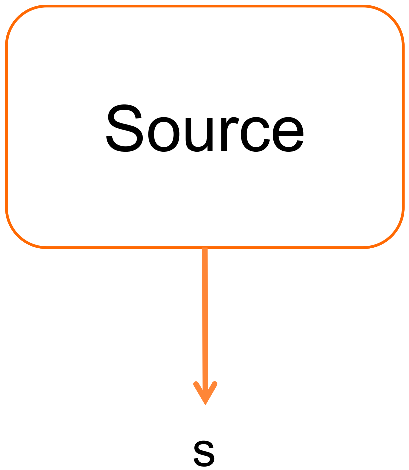
DATA COMPRESSION



Goal: Construct an encoder and decoder such that we can recover the information $\hat{s} = s$

DATA COMPRESSION

- How would you compress the following source?



$$\mathcal{A}_S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$P_S = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0\}$$

Raw bit content

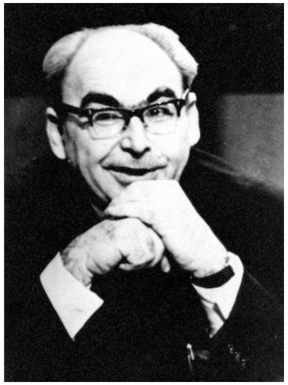
- Measures the number of possible outcomes

$$H_0(S) = \log |\mathcal{A}_S|$$

Same as Shannon entropy if distribution is uniform

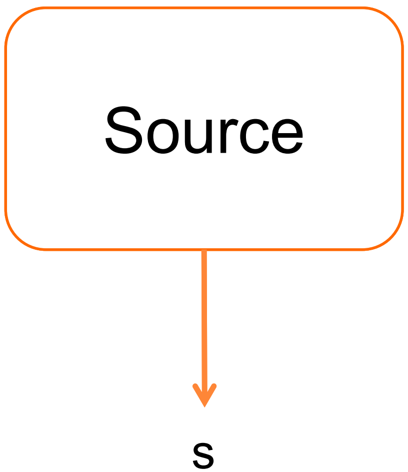
$$|\mathcal{A}_S| = |\{s|P(s) > 0\}|$$

- Also called Rényi entropy of order 0 (Shannon's entropy is the limit of Rényi entropy of order α when $\alpha \rightarrow 1$)



DATA COMPRESSION

- How would you compress the following source?



$$\mathcal{A}_S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$P_S = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$$

Lossy and Lossless compression

- Lossy
  - maps some source symbols to the same encoding
  - failure probability δ
- Lossless compression
  - maps all possible source symbols to distinct encodings
  - works without failure probability

LOSSLESS COMPRESSION (NEXT LECTURE)

- If we don't allow a failure probability, how can it be that we can compress at all?

- Idea
  - Use short encodings for symbols that are very likely
  - Use long encodings for symbols that are unlikely

- If we compress many times we will achieve high compression rates on average as long encodings are not very frequent

Lossy compression

- Allow for failure probability

$$\mathcal{A}_S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

- Back to our example

$$P_S = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$$

- Encode

Lossy compression

- Allow for failure probability
- Back to our example

$$\mathcal{A}_S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$P_S = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$$

- Encode

$$1 \rightarrow 00$$
$$2 \rightarrow 01$$
$$3 \rightarrow 10$$
$$4 \rightarrow 11$$
$$\text{rest} \rightarrow 00 \leftarrow \text{error! } P(rest) = \frac{4}{64} = \frac{1}{16} = \delta$$

Let's try and formalize our idea

- Smallest sufficient subset $S_\delta \subseteq \mathcal{A}_X$

$$P(x \in S_\delta) \geq 1 - \delta$$

- Essential bit content (up to failure probability δ)

$$H_\delta(X) = \log |S_\delta|$$

- In our example

$$|S_\delta| = 4 \qquad |\mathcal{A}_X| = 8$$

Let's try and formalize our idea

- Sn
- Es
- In our example

Can compress to exactly $H_\delta(X)$ bits with failure probability δ

$$|S_\delta| = 4 \qquad |\mathcal{A}_X| = 8$$

COMPRESSING BLOCKS

- Instead of compressing just once, let's consider compressing a long string of bits $X_j \in \{0,1\}$

$$X^N = X_1, \ldots, X_N$$

- IID, probability of

$$P(X_j = 0) = 0.9$$
$$P(X_j = 1) = 0.1$$

- What do you expect?

Compressing blocks

- Instead of compressing just once, let's consider compressing a long string of bits $X_j \in \{0,1\}$

$$X^N = X_1, \ldots, X_N$$

$$H(X^N) = NH(X)$$

- IID, probability of

$$P(X_j = 0) = 0.9$$

$$P(X_j = 1) = 0.1$$

- What do you expect?

Distributions over strings

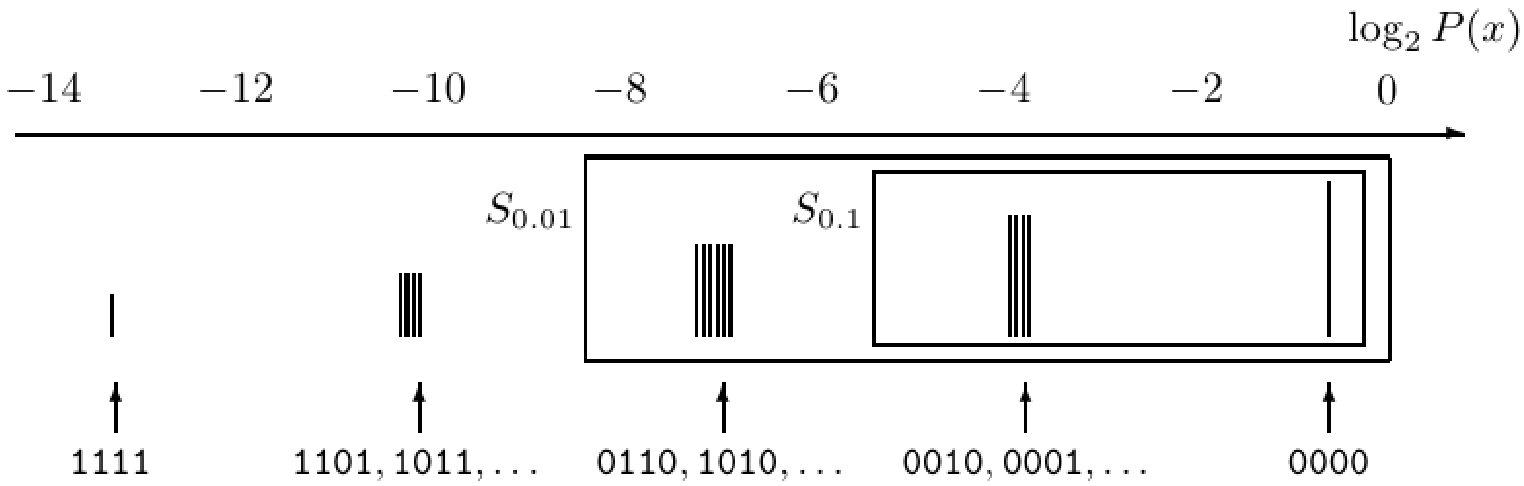
- Probability of a particular string only depends on #1's

$$P(X_j) = P(X_j = 0)^{N-n_1} P(X_j = 1)^{n_1}$$

Distributions over strings

- Probability of a particular string only depends on #1's

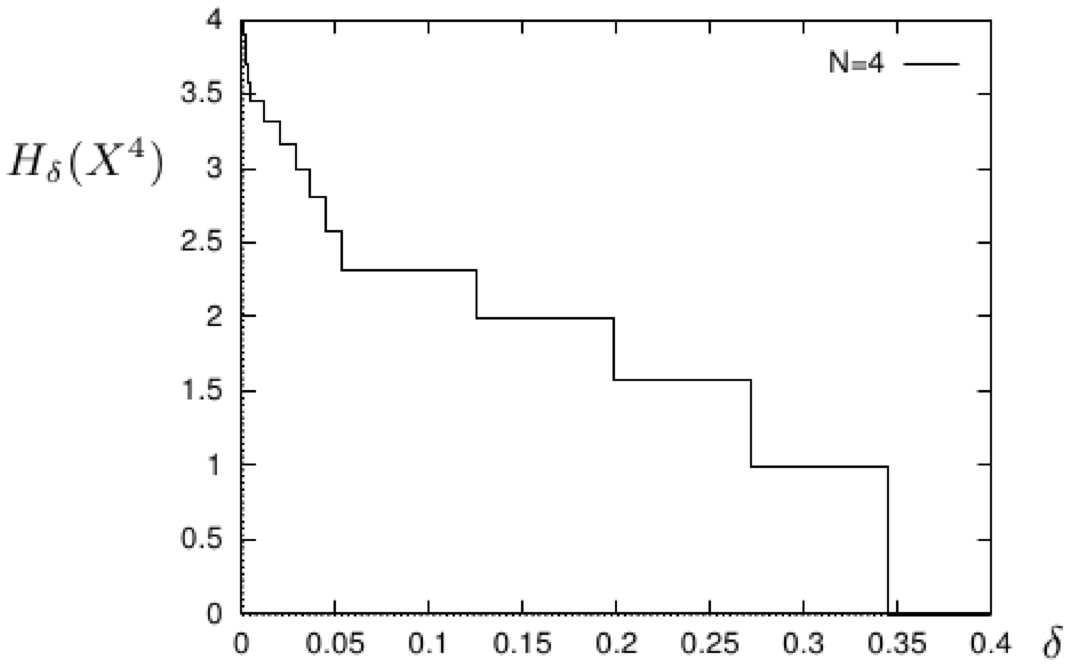
$$P(X_j) = P(X_j = 0)^{N-n_1} P(X_j = 1)^{n_1}$$



ESSENTIAL BIT CONTENT

- Depends on failure probability

$$P(x \in S_\delta) \geq 1 - \delta$$
$$H_\delta(X) = \log |S_\delta|$$



… BUT HOW SIGNIFICANT IS THIS DEPENDENCE?

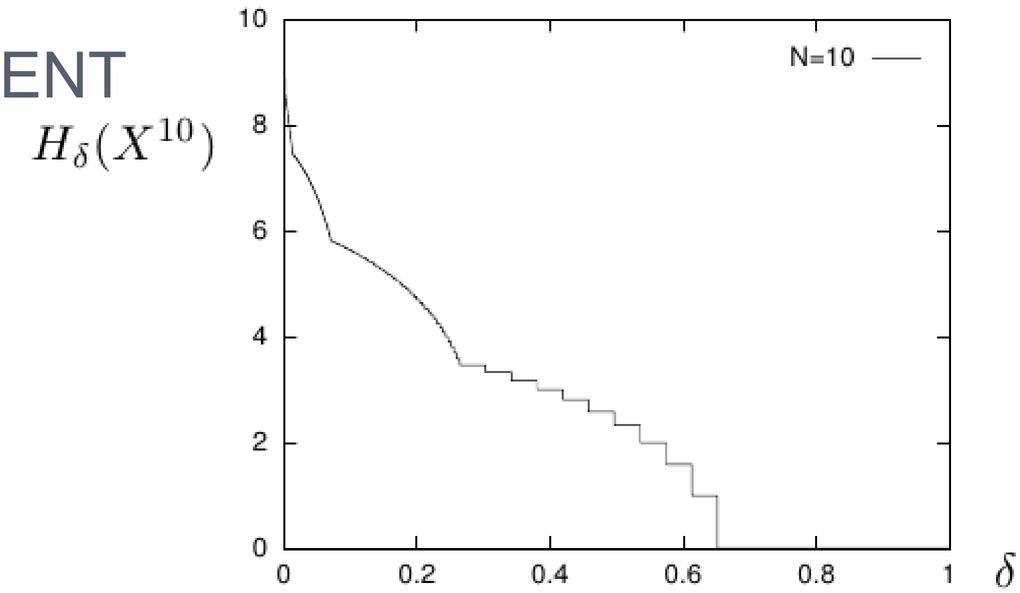
- What do you expect?

Essential bit content

- When N gets large…

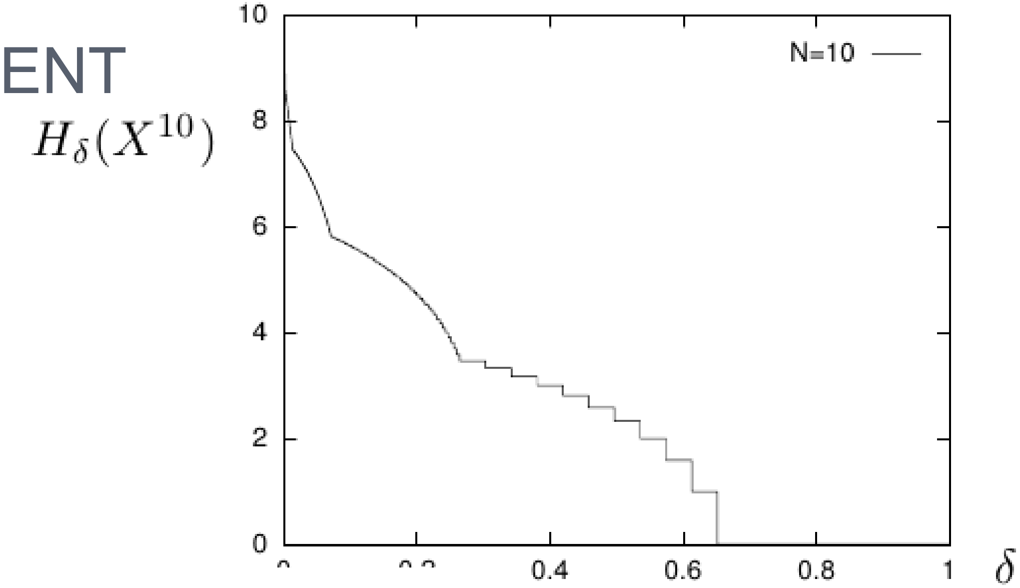
Essential bit content

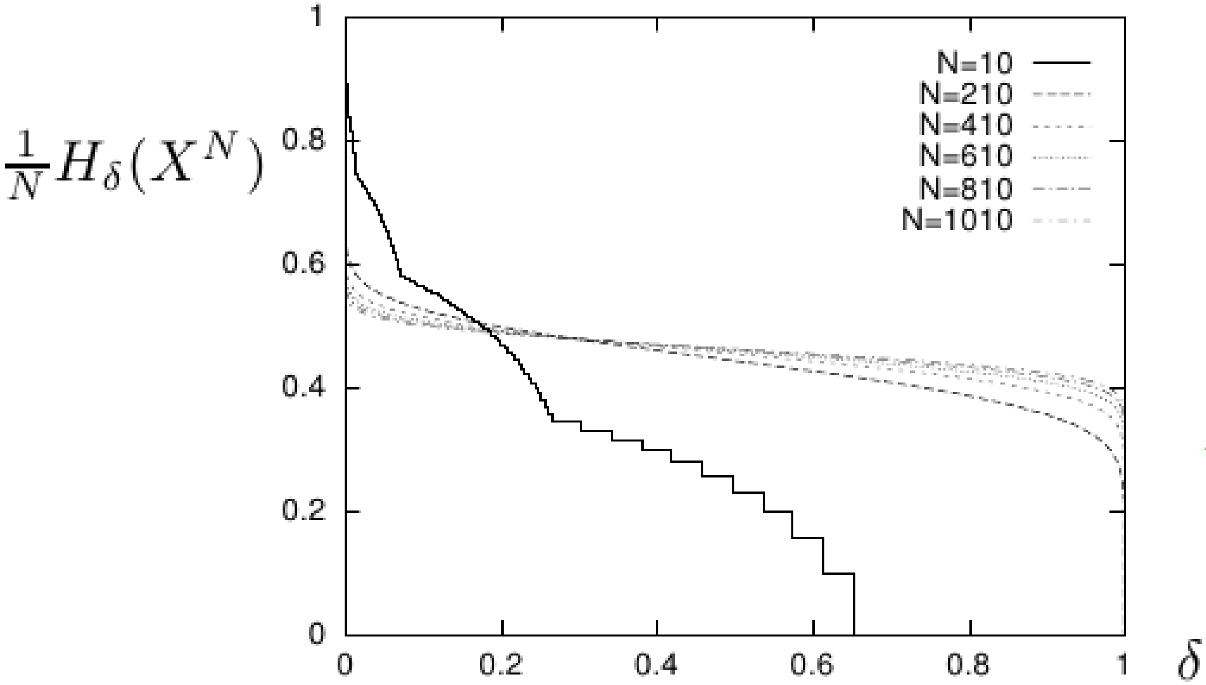
- When N gets large…



ESSENTIAL BIT CONTENT

- When N gets large…





Actual failure probability matters less and less

Shannon's source coding theorem

Let X be a random variable with entropy $H(X)$.

For any $\epsilon > 0$ and $0 < \delta < 1$ there exists

a positive integer N_0 such that for all $N > N_0$

$$\left| \frac{1}{N} H_\delta(X^N) - H(X) \right| \leq \epsilon$$

$$H(X^N) = NH(X)$$

Shannon's source coding theorem

- What does this mean?
- Remember that $H_\delta(X^N)$ tells us how well we can compress with some failure probability
- Informal version

  "N iid random variables each with entropy H(X) can be compressed into roughly N H(X) bits. Conversely if they are compressed further it is virtually certain that information is lost."

We need to prove two things

- There exists a way to compress to $H(X^N) = NH(X)$ bits

$$\frac{1}{N} H_\delta(X^N) < H(X) + \epsilon$$

- If we compress further information is lost

$$\frac{1}{N} H_\delta(X^N) > H(X) - \epsilon$$

How can we determine the sufficient subset?

- In general, hard to compute size of the best one…
- But can approximate it closely with the “typical subset”
- Information content of a “typical string”

$$P(x^N) = P(X=0)^{n_0} P(X=1)^{N-n_0} = p_0^{p_0 N} p_1^{p_1 N}$$

$$-\log P(x^N) \approx N\left(p_0 \log \frac{1}{p_0} + p_1 \log \frac{1}{p_1}\right) = NH(X)$$

Typical subset

- Typical subset, can use entropy to characterize typical strings

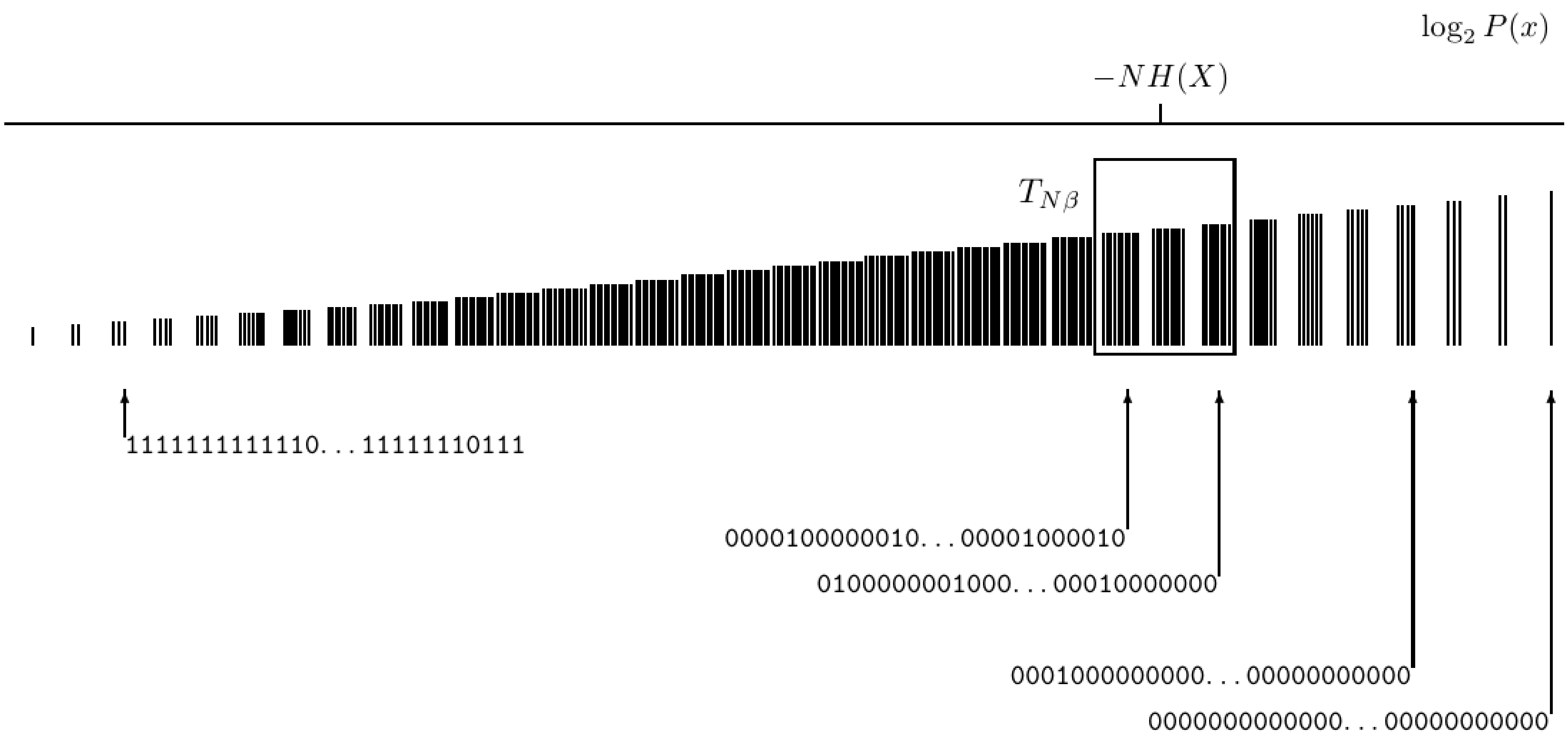
$$T_{N\beta} = \left\{ x^N \in \{0,1\}^N : \left| \frac{1}{N} \log \frac{1}{P(x^N)} - H(x) \right| < \beta \right\}$$

- Compression idea
  - Encode strings in the typical subset
  - Throw the rest away (failure)
  - Because the probability of being in the typical subset if overwhelmingly large the overall failure is very small

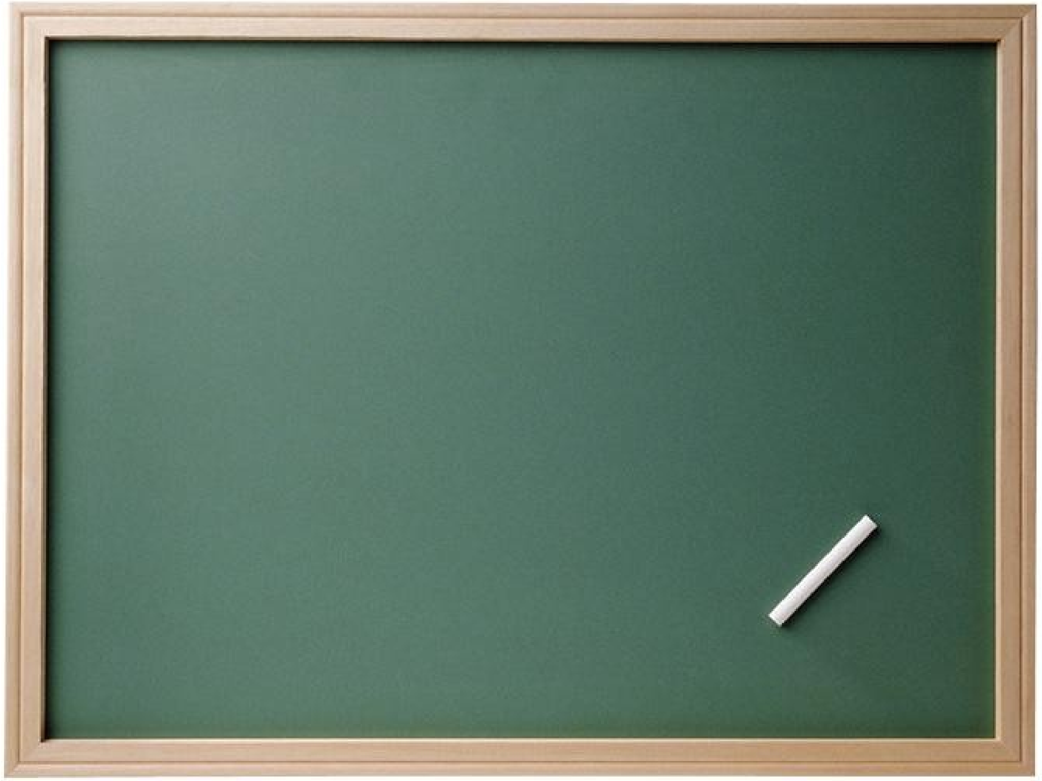
TYPICAL SUBSET FOR OUR EXAMPLE

- String chosen IID with probabilities

$$P(X_j = 0) = 0.9$$
$$P(X_j = 1) = 0.1$$



LET'S PROVE SHANNON'S THEOREM



We need to prove two things

- There exists a way to compress to $H(X^N) = NH(X)$ bits

$$\frac{1}{N} H_\delta(X^N) < H(X) + \epsilon$$

- If we compress further information is lost

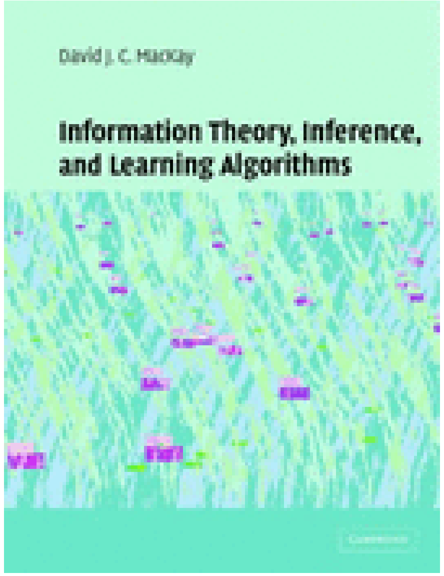
$$\frac{1}{N} H_\delta(X^N) > H(X) - \epsilon$$

Let's summarize

- The entropy of the source tells us exactly how well we can compress it.
- Our compression scheme was
  - Lossy (but error could be arbitrarily small)
  - Fixed block length
    - All elements in our typical subset were assigned an equally long bit string as a codeword
- Next week: lossless compression (symbol codes)

Reading for this lecture

- Chapter 4 in the book



Information Theory, Inference and Learning Algorithms
by David J. C. MacKay
Cambridge University Press, 2003

- Homework due by the beginning of next lecture, hand out in person at next lecture or upload to IVLE Workbin