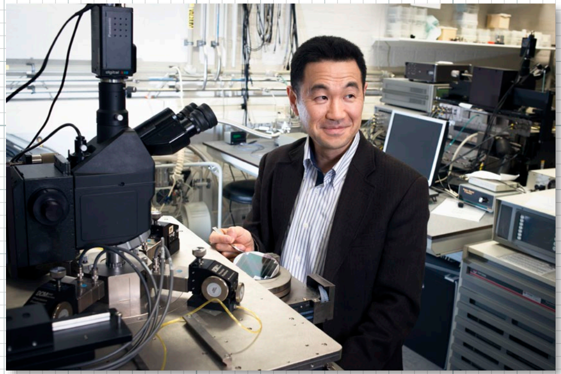
TN3125
Information and Computation

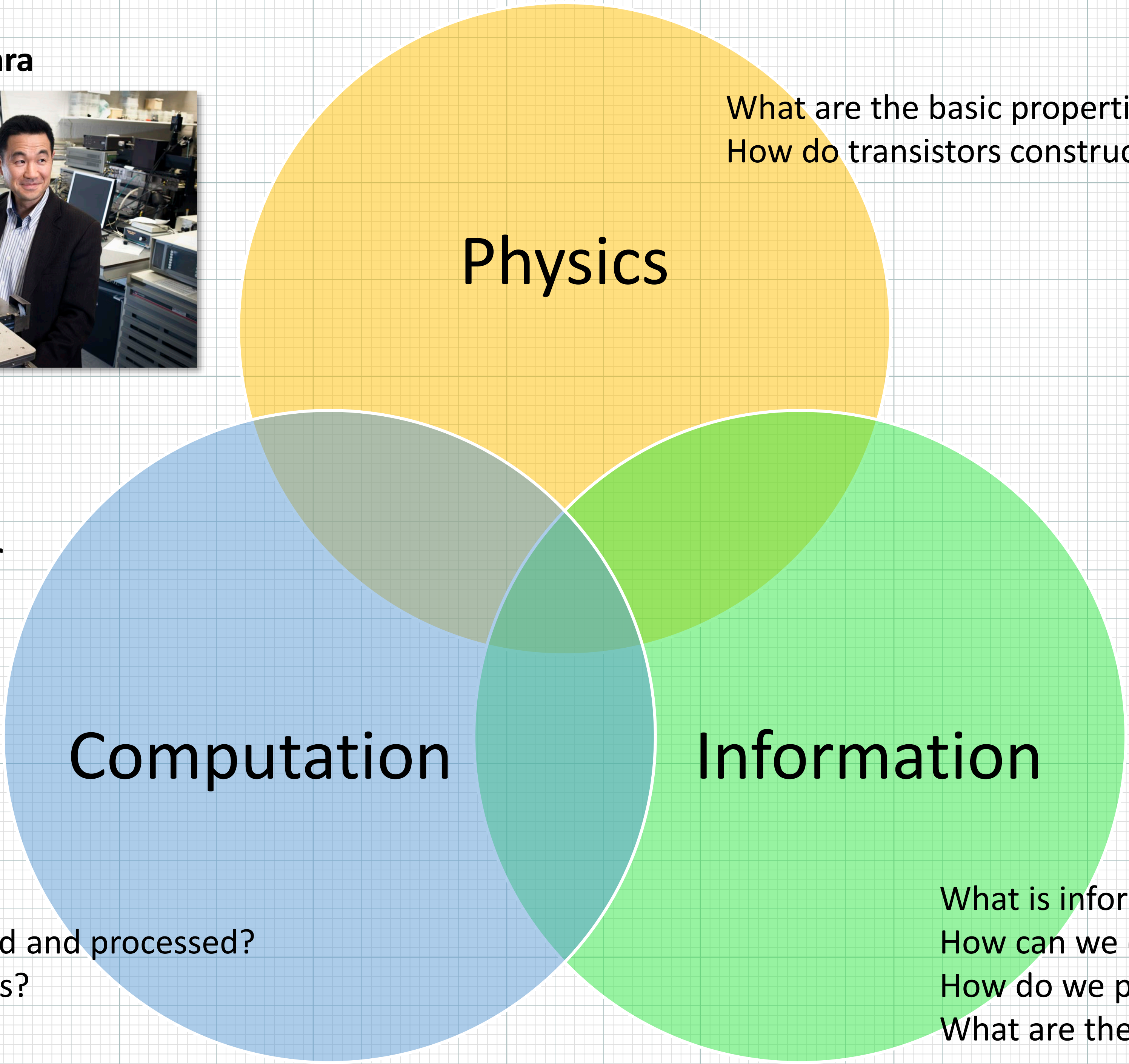
Lecture 1

1A - Introduction

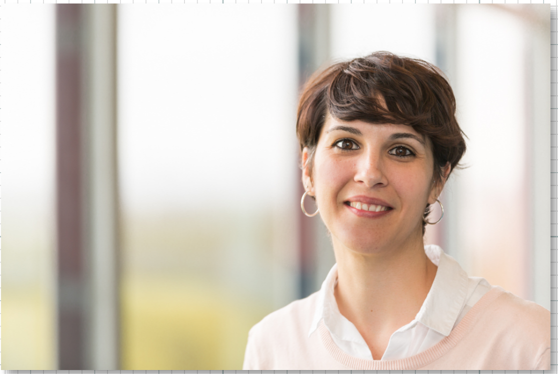
Ryoichi Ishihara



What are the basic properties of hardware?
How do transistors constructing all gates work?



Carmina García Almudever



How do computers work?
How is information encoded and processed?
How to build digital systems?

David Elkouss

What is information?
How can we quantify it?
How do we protect from noise?
What are the fundamental limits?

minor QSQI 2019-2020	11:45-12:30	lunch	13:45-14:30	14:45-15:30	15:45-16:30	16:45-17:30
Monday 2 September						
Tuesday 3 September						
Wednesday 4 September			Information and Computation		Information and Computation	
Thursday 5 September						
Friday 6 September			Information and Computation		Information and Computation	
Monday 9 September						
Tuesday 10 September						
Wednesday 11 September	**Inf. Comp. TEST**		Information and Computation		Information and Computation	
Thursday 12 September						
Friday 13 September			Information and Computation		Information and Computation	
Monday 16 September						
Tuesday 17 September						
Wednesday 18 September	**Inf. Comp. TEST**		Information and Computation		Information and Computation	
Thursday 19 September						
Friday 20 September			Information and Computation		Information and Computation	
Monday 23 September						
Tuesday 24 September						
Wednesday 25 September	**Inf. Comp. TEST**		Information and Computation		Information and Computation	
Thursday 26 September						
Friday 27 September			Information and Computation		Information and Computation	
Monday 30 September						
Tuesday 1 October						
Wednesday 2 October	**Inf. Comp. TEST**					
Thursday 3 October						
Friday 4 October						

Computation

Information

Physics of I&C

Grading scheme

- Option 1: average of the weekly minitests

- Option 2: retake (17^{th} December) grade

Resources

- *Feynman Lectures on Computation*. Richard Feynman.

- **Computation**
  - *Digital design and computer architecture*. D.M. Harris, S.L. Harris. Chapters 1, 2, 3 and 6 (6.1, 6.2, 6.3)
  - Slides
- **Information**
  - *Elements of information theory*. T.M. Cover, J.A. Thomas. Chapters 2, 5 and 7
  - *Lecture notes*
- **Physics**
  - *Semiconductor physics and devices.* D. Neaman. Chapters 11 and 12
  - Slides

TN3125
Information and Computation

Lecture 1

1B - Introduction to information theory

The abstract communications model

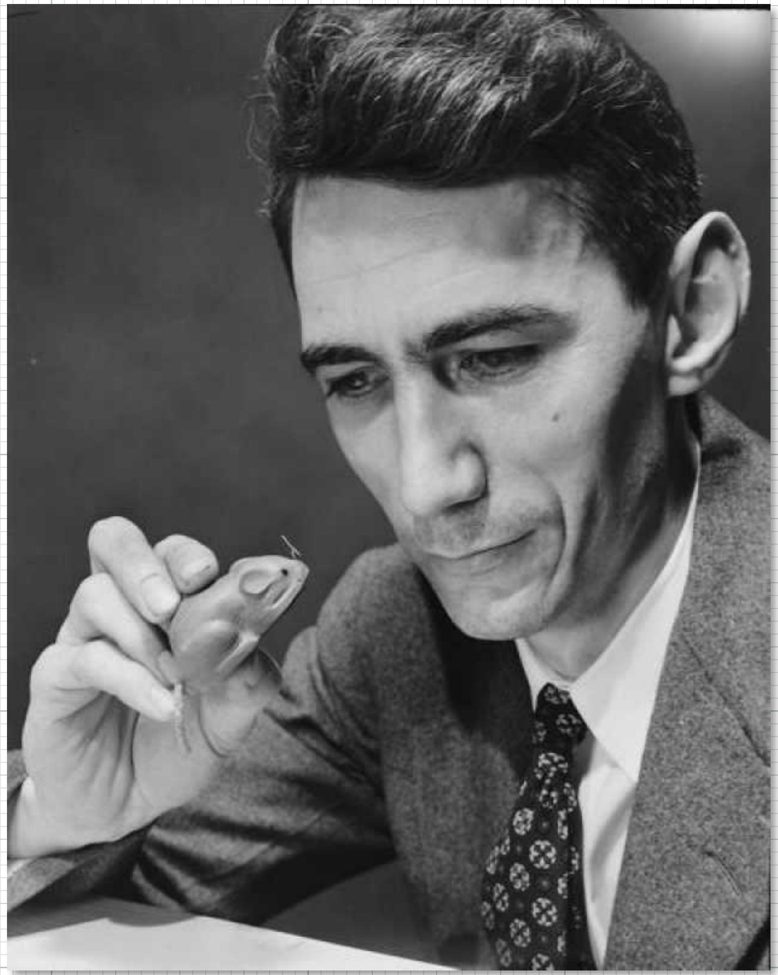
The fundamental problem of communication is that of reliably transmitting information over a channel

Examples:

VOICE → EARS

EARS → BRAIN

ANTENA → SATELLITE



A mathematical theory of communications.
Claude Shannon. Bell systems technical journal (1948)

What is information?

0

1

Vectors by www.Vecteezy.com

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. You receive a 0. Does the message contain information?

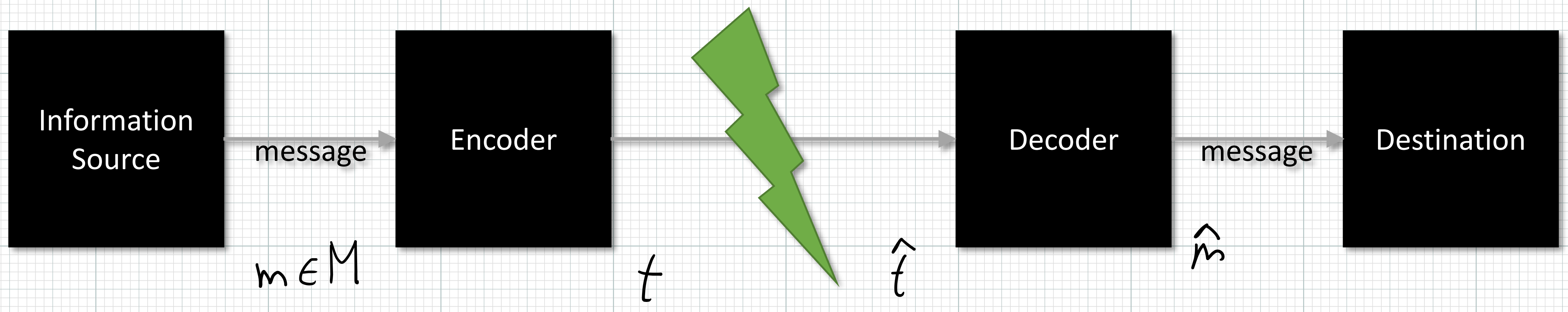
	Days with no rain	Days with rain
Summer	182	1
Winter	30	152

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. You receive a 0. Does the message contain information?

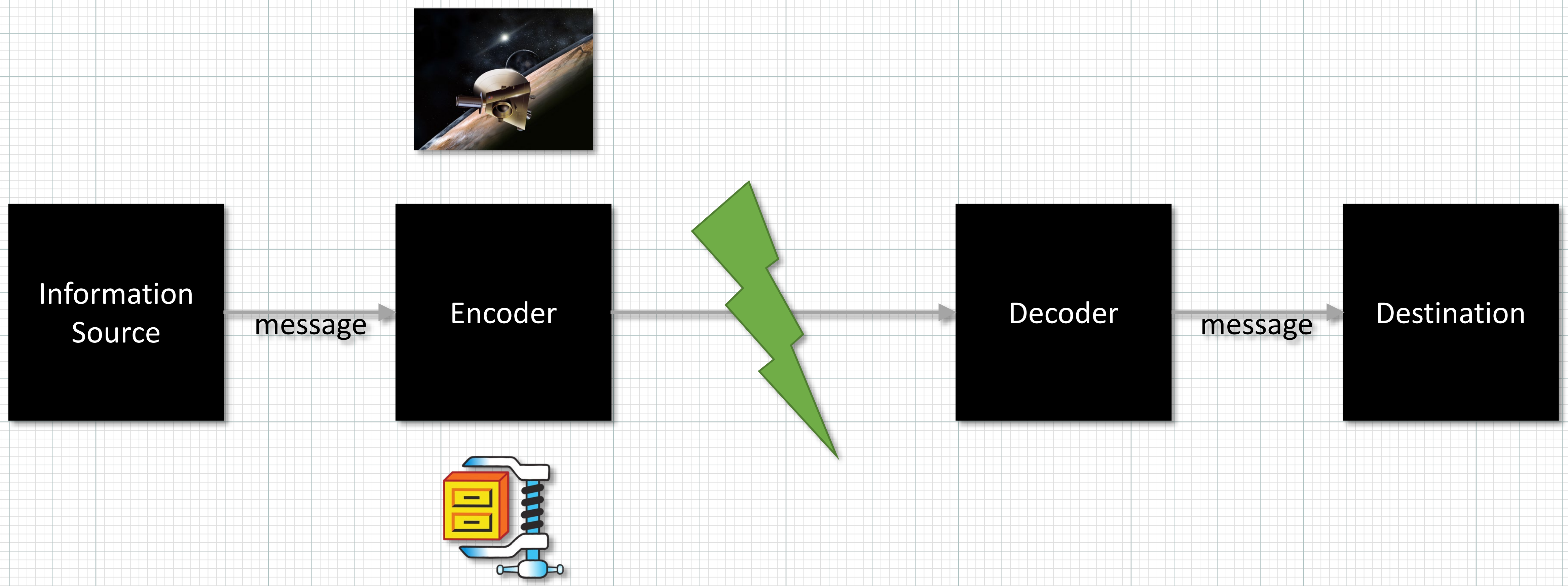
Exercise 1.3 - Finally, let us assume that you live in the Netherlands but (boldly) also that you are aware of the current season. You receive a 0. Does the message 0 carry the same information in summer and in winter?

The abstract communications model



$$m = \hat{m}$$

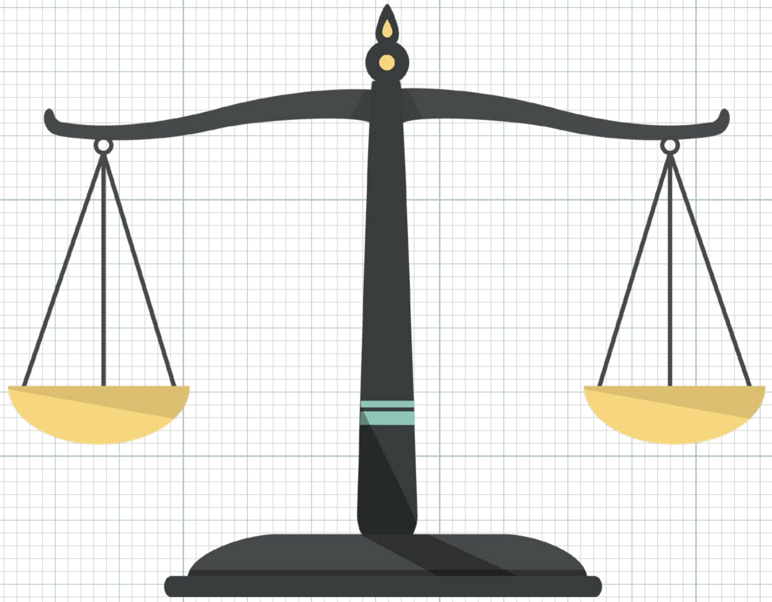
The abstract communications model



Lecture 1: Learning goals

- Quantify the information of an event
- Quantify the average information of a probability distribution
- Quantify conditional information
- Derive basic properties of the entropy function

Exercise. There are 12 coins, one of which is a counterfeit. The counterfeit is either lighter of heavier than normal coins but not of the same weight. You are given a two plate scale where you can compare weights.



- Devise a scheme that detects the counterfeit coin when it is heavier
- Devise a scheme that detects the counterfeit coin in the general case

$1^{\pm}$
$2^{\pm}$
$3^{\pm}$
$4^{\pm}$
$5^{\pm}$
$6^{\pm}$
$7^{\pm}$
$8^{\pm}$
$9^{\pm}$
$10^{\pm}$
$11^{\pm}$
$12^{\pm}$

$$\frac{1234}{5678}$$

$+$

$1^{+}\ 5^{-}$
$2^{+}\ 6^{-}$
$3^{+}\ 7^{-}$
$4^{+}\ 8^{-}$

$$\frac{125}{346}$$

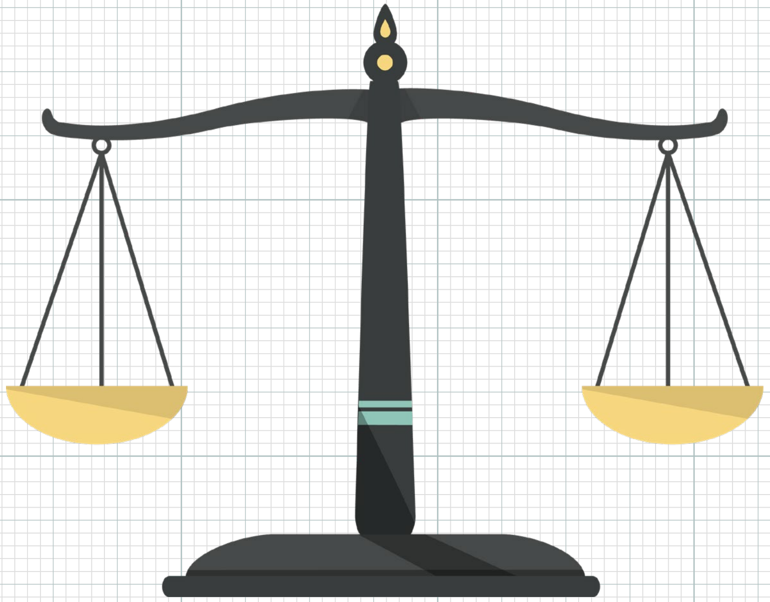
1^{+}
2^{+}
6^{-}

7^{-}
8^{-}

$9^{\pm}$
$10^{\pm}$
$11^{\pm}$
$12^{\pm}$

$$\frac{9\,10}{1\,2}$$

Exercise. There are 12 coins, one of which is a counterfeit. The counterfeit is either lighter of heavier than normal coins but not of the same weight. You are given a two plate scale where you can compare weights.



- Devise a scheme that detects the counterfeit coin when it is heavier
- Devise a scheme that detects the counterfeit coin in the general case
- Is it possible to detect the counterfeit coin with three weighings? And with two?

TN3125
Information and Computation

Lecture 1

2 – A probability refresher and an information measure

A probability refresher

- **Definition.** A probability distribution function is a function $p: X \mapsto [0,1]$ such that $\sum_{x \in X} p(x) = 1$
- **Example:** $X = \{tails, heads\}$, $p(tails) = 0.3$
  $p(heads) = 0.7$
- **Example (uniform distribution):**
  $|X| = n$ $p(x) = \frac{1}{n}$

A probability refresher

- We can extend the definition of probability distribution to sets. Given $S \subseteq X$

$$p(S) = \sum_{s \in S} p(s)$$

- An event is a subset of X. The algebra of sets translates to logic of events: $p(S \cup T) = p(S \text{ or } T)$, $p(S \cap T) = p(S \text{ and } T)$, etc.
- **Example:** $X = \{1, 2, 3\}$

$S = \{1, 2\}$

$T = \{1, 3\}$

$p(S \cap T) = \frac{1}{3}$

$p(S \cup T) = 1$

A probability refresher

- **Definition.** An ensemble X is a tuple (A_X, p_X)
- **Definition.** Given an ensemble X and two events a, b:

$$p_X(a|b) = \frac{p_X(a \text{ and } b)}{p_X(b)}$$

- **Definition.** Given an ensemble X and two events a, b, they are independent if

$$p_X(a \text{ and } b) = p_X(a)p_X(b)$$

A probability refresher

- **Exercise.** Let X be an ensemble modelling two fair coins. I.e.:

$$
\begin{aligned}
p_X(\{\text{tails,tails}\}) &= p_X(\{\text{tails,heads}\}) \\
&= p_X(\{\text{heads,tails}\}) \\
&= p_X(\{\text{heads,heads}\}) \\
&= 1/4
\end{aligned}
$$

- Identify two events a, b that are independent and verify that $p_X(a \text{ and } b) = p_X(a)p_X(b)$.

a = {heads, tails}
{heads, heads}

b = {tail, heads}
∪ {heads, heads}

A probability refresher

- **Definition.** A random variable V on the ensemble X is a function $V: A_X \mapsto A_V$, where A_V is a subset of the reals. V induces an ensemble (A_V, p_V) where p_V is given by:

$$p_V(v) = \sum_{x \in X : V(x) = v} p_X(x)$$

- **Definition.** The mean of a random variable V is given by:

$$E[V] = \sum_{v \in V} v \cdot p_V(v)$$

Axiomatic derivation of an information measure

- Let X be an ensemble and h an information measure, where $h: A_X \mapsto \mathbb{R}$
- We are going to impose a series of properties on h
- Our goal will be to find a set of functions that verifies the conditions

Axiomatic derivation of an information measure

- If a, b are independent events:

$$h(a \text{ and } b) = h(a) + h(b)$$

- For any event a:

$$h(a) \geq 0$$

- For all events a, b:

$$\text{If } p(a) > p(b) \text{ then } h(a) < h(b)$$

- The measure should be continuous

The entropy function

- Given an ensemble X, the function

$$h(x) = -\log_\lambda(p_X(x))$$

with $\lambda > 1$ satisfies all the requirements! **(check)**

- We choose $\lambda = 2$ and let $\log = \log_2$
- The function h is a random variable that only depends on p_X
- Example: $p(\text{tails}) = \frac{1}{2}$

$$-\log \frac{1}{2} = 1$$

$$-\log 2^{-n}$$

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. You receive a 0. Does the message contain information?

Exercise 1.3 - Finally, let us assume that you live in the Netherlands but (boldly) also that you are aware of the current season. You receive a 0. Does the message 0 carry the same information in summer and in winter?

Entropy rate

- **Definition.** The entropy (rate) of an ensemble X is
$$H(X) = -\sum_{x\in X} p_X(x)\log p_X\,(x)$$
where we adopt the convention $0\log 0 = 0$.
- What is the relation between H(X) and h?

Exercises

- What is the entropy of a fair coin?

- What is the entropy of a coin that always yields tails?

- Can entropy be smaller than 0?

Exercises

Let X be an ensemble with $H(X) > 0$ and let $Y = f(X)$ $\forall x \in X \quad f(x) = x$

- Give one function such that $H(X) = H(Y)$
- Give one function such that $H(Y) = 0$
- Give one function such that $H(X) > H(Y) > 0$
- Give one function such that $H(X) < H(Y)$

$$H(X) - H(Y) =$$

$$a^{x+y} = a^x a^y$$

$$\sum_x -p_x \log(p_x) - \sum_y -p_y \log(p_y) =$$

$$= -p_{x_1} \log p_{x_1} - p_{x_2} \log p_{x_2} - (-p_{y_1} \log p_{y_1})$$

$$= \log \frac{1}{(p_{x_1})^{p_{x_1}}} + \log \frac{1}{(p_{x_2})^{p_{x_2}}} + \log (p_{y_1})^{p_{y_1}}$$

$$= \log \frac{(p_{y_1})^{p_{y_1}}}{(p_{x_1})^{p_{x_1}} (p_{x_2})^{p_{x_2}}} = \log \frac{(p_{x_1} + p_{x_1})^{p_{x_1} + p_{x_2}}}{(p_{x_1})^{p_{x_1}} (p_{x_1})^{p_{x_2}}}$$

2

$$= \log \frac{(P_{x_1} + P_{x_1})^{P_{x_1} + P_{x_2}}}{(P_{x_1})^{P_{x_1}} (P_{x_1})^{P_{x_2}}} = \log \frac{(P_{x_1} + P_{x_2})^{P_{x_1}}}{(P_{x_1})^{P_{x_1}}} + \log \frac{(P_{x_1} + P_{x_2})^{P_{x_2}}}{P_{x_2}^{P_{x_2}}}$$

$$= P_{x_1} \log \frac{P_{x_1} + P_{x_2}}{P_{x_1}} + P_{x_2} \log \frac{P_{x_1} + P_{x_2}}{P_{x_2}}$$

TN3125
Information and Computation

Lecture 1

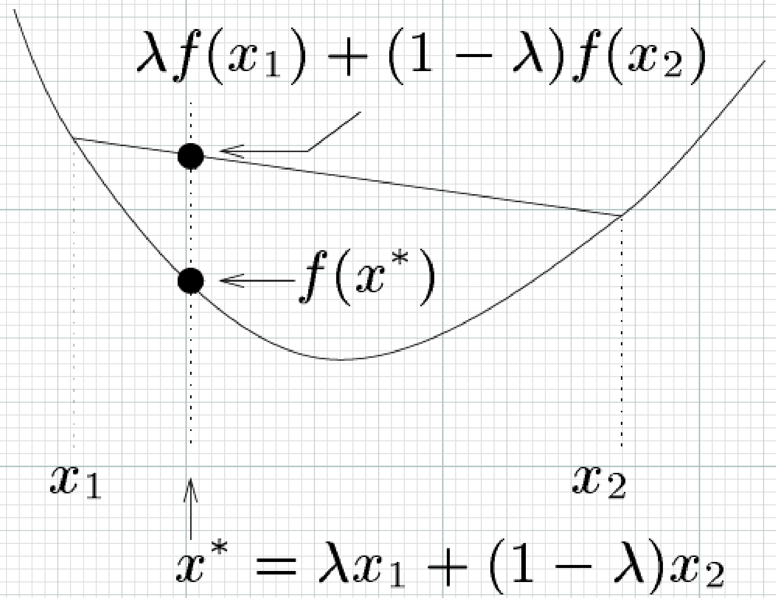
3 – Properties of entropy

Concave functions

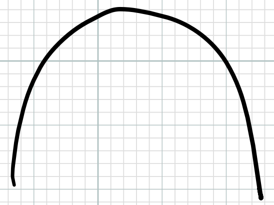
- **Definition.** A function f is concave if in the interval (a, b) if for all x_1, x_2 in the interval and $\lambda \in [0,1]$:

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2)$$

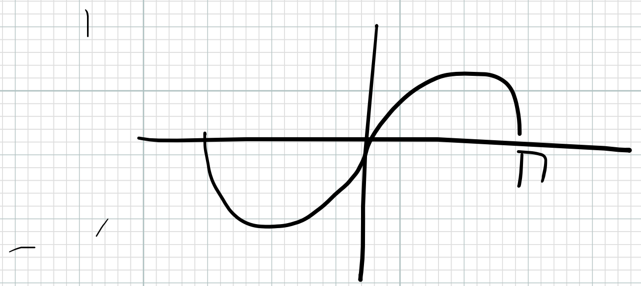
- **Lemma.** A function f is concave in the interval (a, b) if it is twice differentiable and $f'' \leq 0$ in the interval



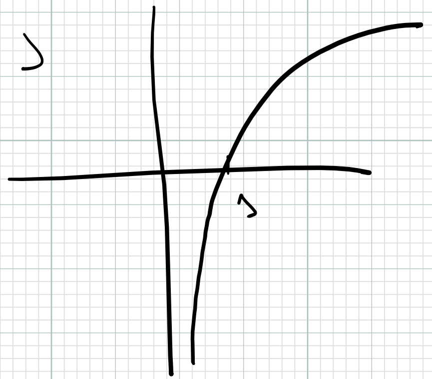
$-x^2$

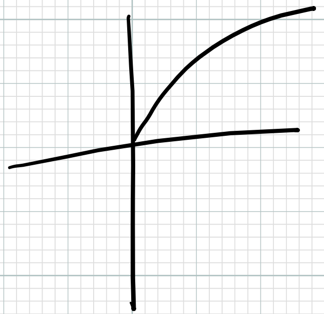


$\sin(x),\ x \in (0, \pi)$



log





Exercise

- Show that $f(x) = -x \log x$ is concave in the interval $(0,1)$

$$f'(x) = -\log x - 1$$

$$f''(x) = -\frac{1}{x}$$

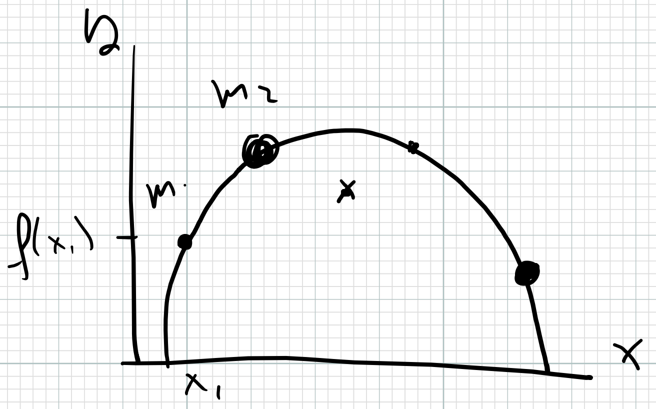
Jensen's theorem

- Let f be a concave function in the interval (a, b). Then for any set of points $x_1, \dots, x_n$ in the interval and for $p_1, \dots, p_n$ non-negative and adding up to one:

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{i=1}^{n} p_i f(x_i)$$

Exercise

- Show that the center of masses of a set of masses placed over a concave curve, that is at locations $(x_i, f(x_i))$, lies below the curve.



$$c_x = \frac{\sum_i x_i \cdot m_i}{\sum_i m_i} = \sum_i x_i \frac{m_i}{M}$$

$$c_y = \sum_i f(x_i) \cdot \frac{m_i}{M}$$

$$f(c_x) = f\left(\sum_i x_i \frac{m_i}{M}\right) \geq \sum_i \frac{m_i}{M} f(x_i) = c_y$$

Exercise

Find the entropy of the uniform distribution on n elements

$$H(x) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = -\log \frac{1}{n} = \log n$$

Exercise

- Show that the distribution that maximizes entropy for any alphabet size is the uniform distribution. (Hint: use Jensen’s inequality and the concavity of the log function).

Exercise (concavity of entropy)

- Let X_1, X_2 be two independent random variables modelling two different but otherwise indistinguishable bent coins.
- Let Y be a binary random variable that models a process where we throw X_1 with probability t and X_2 with probability $1 - t$.
- Show that

$$H(Y) \geq tH(X_1) + (1 - t)H(X_2)$$

$f(x) = -x \log x$ is concave

$$y_i = t x_i^1 + (1-t) x_i^2$$

$$H(y) = \sum_i -y_i \log y_i \geq t \sum_i -x_i^1 \log x_i^1 + (1-t) \sum_i -x_i^2 \log x_i^2$$

$$= t H(x_1) + (1-t) H(x_2)$$

Exercise (telescopic property)

- Let X_1, X_2 be two independent random variables modelling two different perfectly distinguishable bent coins.
- Let Y be a binary random variable that models a process where we throw X_1 with probability t and X_2 with probability $1-t$.
- Show that

$$H(Y) = H(t, 1-t) + tH(X_1) + (1-t)H(X_2)$$

Information and Computation

Lecture 1

Joint entropy, conditional entropy and mutual information

A probability refresher

- **Definition.** A joint ensemble XY is a tuple $(\mathrm{A}_X \times A_Y, p_{XY})$ with

$$p_X(x\ ) = \sum_{y \in A_Y} p_{XY}(x, y)$$

$$p_Y(y\ ) = \sum_{x \in A_Y} p_{XY}(x, y)$$

$$p_{XY}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

Joint entropy

- **Definition.** The entropy (rate) of an ensemble XY is

$$H(XY) = -\sum_{x \in X, y \in Y} p_{XY}(x,y) \log p_{XY}(x,y)$$

- **Example.**

$$p(\text{tails}, \text{tails}) = \qquad = \frac{1}{4}$$

$$H(XY) = -\sum \frac{1}{4} \log \frac{1}{4}$$

$$= \frac{1}{4}(2 + 2 + 2 + 2) = 2$$

Conditional entropy

- **Definition**. Conditional information,

$$h(x|y) = -\log(p(x|y))$$

- **Definition**. Conditional entropy given event y

$$H(X|y) = -\sum_{x \in A_X} p(x|y) \log(p(x|y))$$

- **Definition**. Conditional entropy

$$H(X|Y) = \sum_{y} p_Y(y) H(X|y)$$

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Summer	182	1
Winter	30	152

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. You receive a 0. Does the message contain information?

Exercise 1.3 - Finally, let us assume that you live in the Netherlands but (boldly) also that you are aware of the current season. You receive a 0. Does the message 0 carry the same information in summer and in winter?

	Days with no rain	Days with rain
Summer	182	1
Winter	30	152

- Compute $H(\text{Season}), H(\text{Weather})$
- Compute $H(\text{Season,Weather})$
- Compute $H(\text{Season}|\text{Weather=rain})$

P	NO R	RAIN
Sum	$\frac{182}{365}$	$\frac{1}{365}$
Wint	$\frac{30}{365}$	$\frac{152}{365}$

Exercise

- Show that $H(X|Y) \geq 0$

Exercise. Chain rule

- Show that $H(XY) = H(X|Y) + H(Y)$

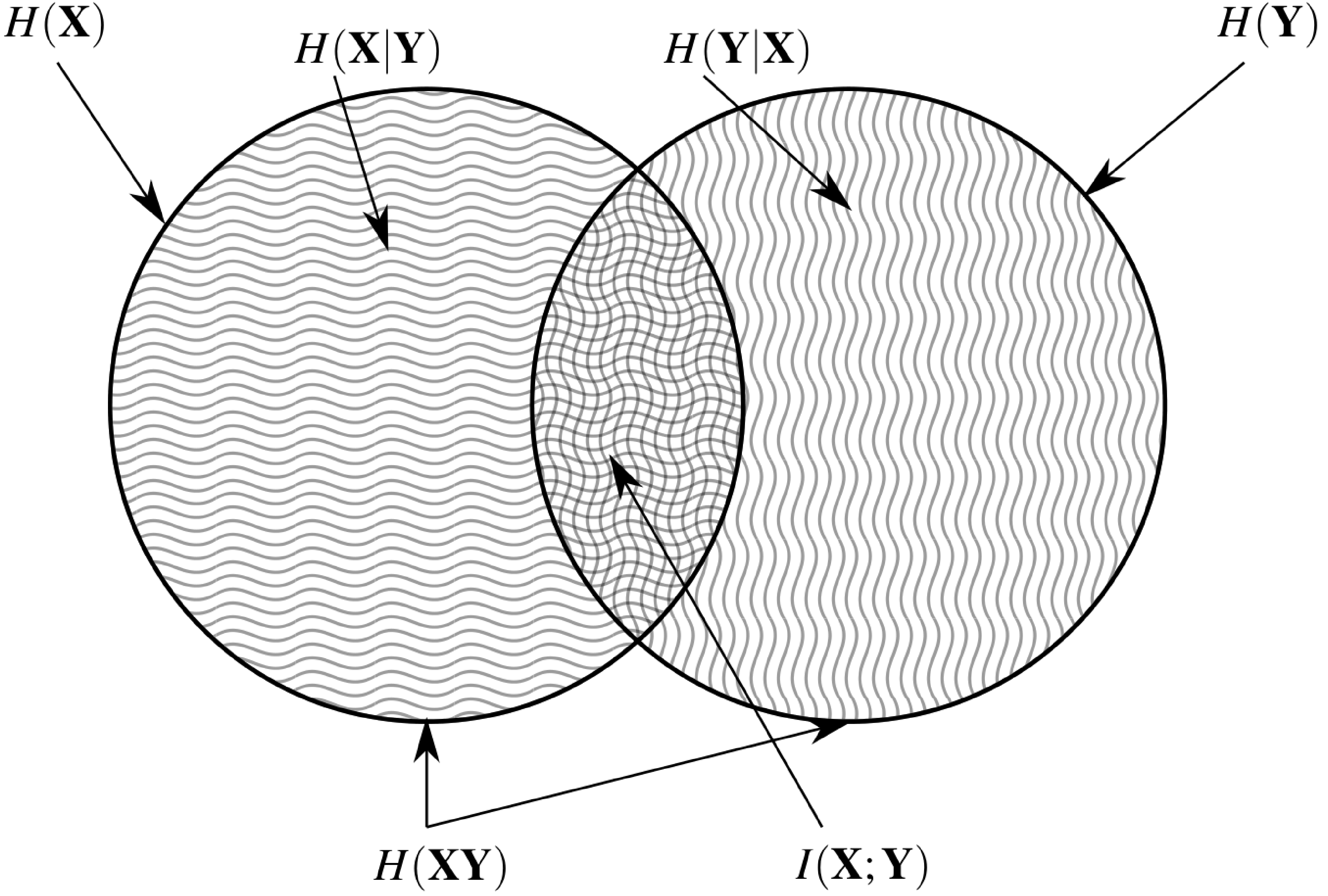
Exercise

- Show that
  - $H(X,Y) \geq H(X)$
  - $H(X) \geq H(X|Y)$ (conditioning always reduces entropy)

Mutual information

- **Definition**. Mutual information

$$I(X;Y) = H(X) + H(Y) - H(XY)$$



$H(\mathbf{X})$
$H(\mathbf{X}|\mathbf{Y})$
$H(\mathbf{Y}|\mathbf{X})$
$H(\mathbf{Y})$
$H(\mathbf{XY})$
$I(\mathbf{X};\mathbf{Y})$

Exercise

- Show that
  - $\min\big(H(X), H(Y)\big) \geq I(X;Y) \geq 0$

Data processing inequality

- Show that $I(X;Y) \geq I(X;f(Y))$

Let $Z = f(Y)$ and let us expand $I(X:YZ)$ in two ways:

1) $$I(X;YZ) = H(X) + H(YZ) - H(XYZ) + H(Y) - H(Y) + H(XY) - H(XY)$$

$$= I(X;Y) + H(YZ) - H(XYZ) - H(Y) + H(XY)$$

$$= I(X;Y) + \underbrace{\cancel{H(Z|Y)}}_{0} - \underbrace{\cancel{H(Z|XY)}}_{0}$$

$$= I(X;Y)$$

2) $I(X:YZ) = I(X:Z) + H(YZ) - H(XYZ) - H(Z) + H(XZ)$

$$= I(X:Z) + H(X|Z) - H(XY|Z)$$

$$= I(X:Z) + \sum_z p(z)\Big(H(X|z) - H(XY|z)\Big)$$

$$= I(X:Z) + \sum_z p(z)\, I(X:Y|z)$$

Hence:

$$I(X:Y) = I(X:Z) + \sum_z p(z)\, I(X:Y|z)$$

$$\geq I(X:Z)$$

You will do great in the exam if you can

- Evaluate entropy, conditional entropy and mutual information
- Derive basic properties of the above

Resources

- Lecture notes
- Slides
- “Elements of information theory”. Cover and Thomas. Chapter 2.1-2.8
- “Information Theory, Inference, and Learning Algorithms”. MacKay. Chapter 2