Lesson 4
Entropy
4-1: Information Measures

Jan Reimann

Math 574, Topics in Logic
Penn State, Spring 2014

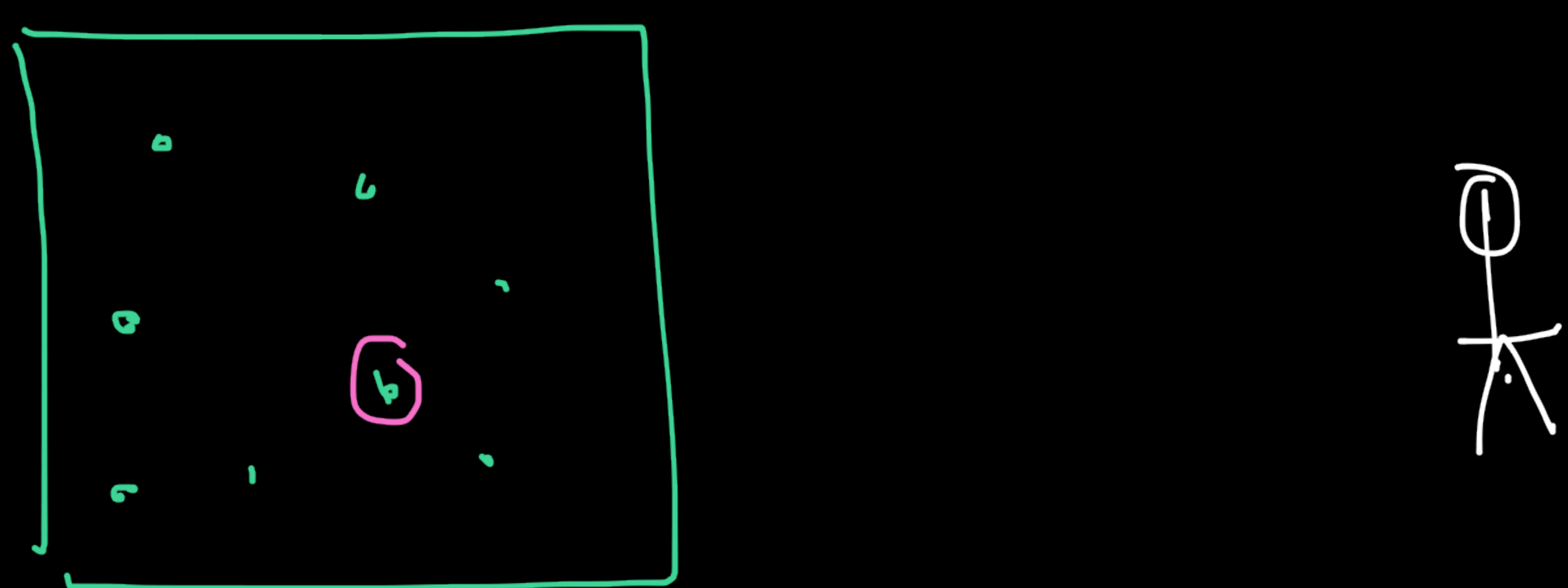
PENNSTATE
Eberly College of Science

Combinatorial Information

Suppose X is a finite, non-empty set. We choose an element in X and want to communicate the information which $x \in X$ we chose to someone else.

Binary channel: We can transmit only strings of 0 and 1.

Question: How many bits are needed to to transmit information about x, if nothing else is known about X but its cardinality?

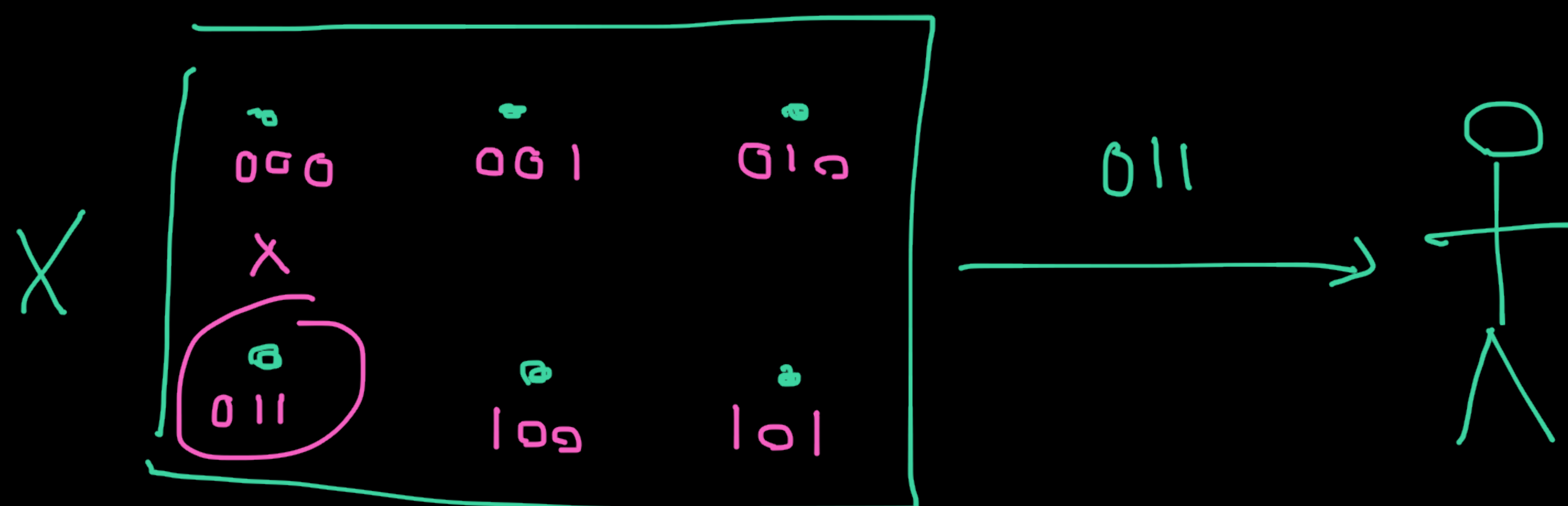


Combinatorial Information

We assign every element in X a fixed-length binary code:

$$c : X \to \{0,1\}^n \quad \text{(one-one)}$$

and transmit $c(x)$.



We need $n = \lceil \log_2 |X| \rceil$ bits.

In this lesson, unless otherwise stated, all logarithms are binary. We write $\log_2 = \log$.

Combinatorial Interpretation

Information = Bits needed to transmit = log(# of possibilities)

$\log |X|$

Probabilistic Information Transfer

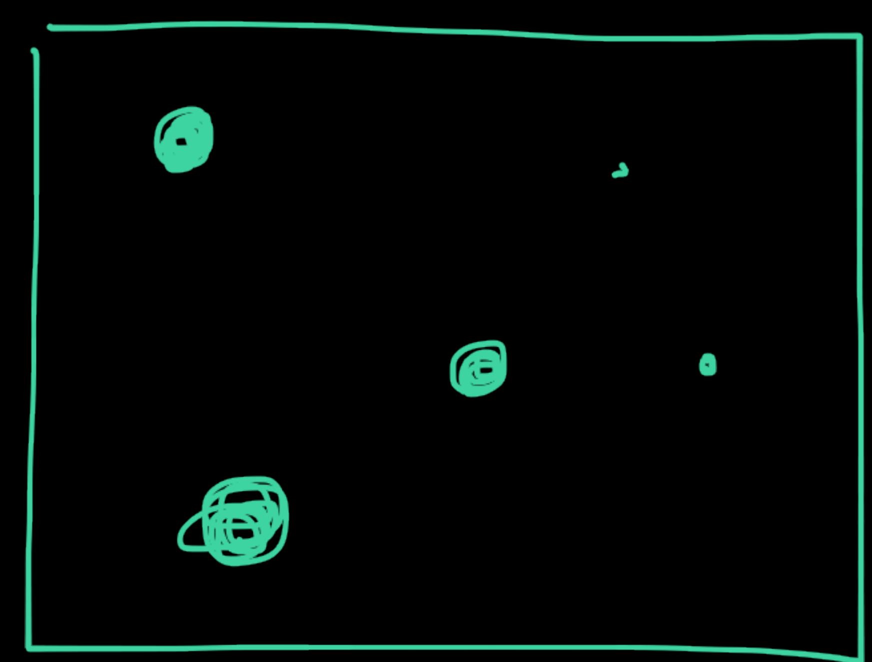
Now assume we have probability measure *P* on *X* (still finite).

We draw $x \in X$ at random according to *P*.

Question: Can we design a code $c : X \to 2^{<\mathbb{N}}$ (not necessarily of fixed length) such that

the expected length of a codeword is minimal?

$\longrightarrow$ Entropy

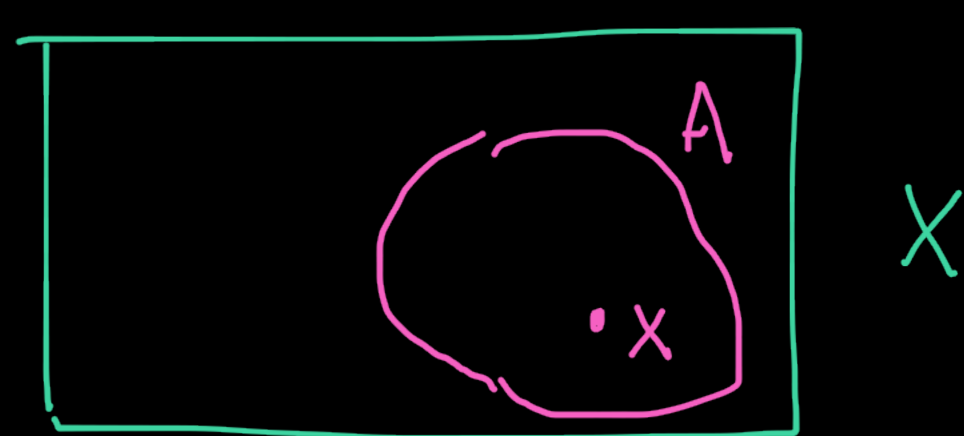


Information Gained = Uncertainty Removed

If, in the combinatorial framework, instead of giving a full description of $x \in X$, we only give a set $A \subseteq X$ with $x \in A$.

How much information about x do we gain from knowing A?

- Clear: the smaller A (compared to X), the more information gained, and the more uncertainty about x removed.
- If $A = X$, we gain no information at all.



Information Functions

We are looking for a function $I : \mathbb{Q}^{\cap}(0,1] \to \mathbb{R}^{\geqslant 0}$. The idea is that $I(|A|/|X|)$ measures the information content of A.

- $I(1) = 0$ (no information if $A = X$)
- I is decreasing.

Observation: $I(|A|/|X|) = -\log(|A|/|X|)$ has these properties

Probabilistic Information

Now let $(X, \mathcal{B}, \mu)$ be a probability space.

$|A| / |X|$

$\downarrow$

$\mu(A)$

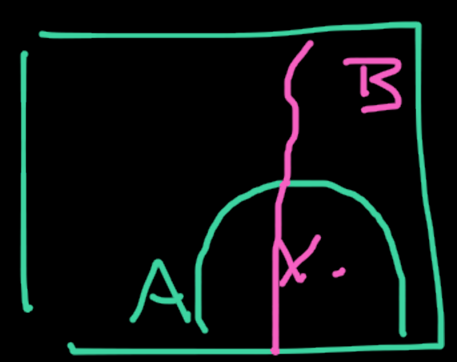
We can generalize the previous approach. We are looking for a function $I : (0, 1] \to \mathbb{R}^{\geqslant 0}$ such that $I(\mu(A))$

(I1) $I(1) = 0$ (no information for almost sure events)

(I2) I is decreasing in $\mu(A)$.

In the probabilistic setting, we have the notion of independence. With respect to information contect, we would want that

information gained through independent events should behave additively,



that means

(I3) whenever $\mu(A \cap B) = \mu(A)\mu(B)$,

$$I(\mu(A \cap B)) = I(\mu(A)) + I(\mu(B))$$

.

Probabilistic Information Function

PROP: If $I : (0,1] \to \mathbb{R}^{\geqslant 0}$ is a continuous function satisfying (I1) - (I3), there must exist a constant $c > 0$ such that

$$I(x) = -c \log(x).$$

Hence $-c \log$ is the only reasonable information function.

entropy = expected gain of information

PENNSTATE Eberly College of Science