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

$a = \{\text{heads, tails}\}$, $\{\text{heads, heads}\}$

$b = \{\text{tail, heads}\} \cup \{\text{heads, heads}\}$

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 $h\colon A_X \mapsto \mathbb{R}$
- 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

Vectors by www.Vecteezy.com

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:

If $p(a) > p(b)$ 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 $Y = f(X)$ $\quad \forall x \in X \quad f(x) = x$

- 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

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

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