Course overview & history

Information theory:
theoretical limitations and potentials of imperfect/noisy systems to store and transmit information

how to measure "information (content)"

Coding theory:
creation of practical en-/decoding schemes

Other areas:
probability theory, statistics & inference, physics, computer science, cryptography

1940:
thought to be impossible to send information at a positive rate with arbitrarily small probability of error (when the transmission device has any nonzero failure rate)

1948: Shannon’s paper
- Arbitrarily small error probability is achievable for all information transmission rates below the “capacity” of the channel. This capacity can be computed and is basically always positive.
- Also showed how to compress data to their minimal size with arbitrarily high recovery probability
since then: explosion of the field in science and engineering; many applications such as

- data compression (ZIP, JPEG, MP3,...)  
- error correction (hard disks, CD,...)  
- channel coding (WLAN, mobile network,...)

classical information and coding theory

1764: Gordan asks whether more (classical) information can be transmitted when exploiting the quantum nature of devices

1970's: upper bound on capacity (devlin, Holevo) "Holevo bound"  
- quantum estimation theory (Helstrom)  
- strong subadditivity of quantum entropy (Lieb, Ruskai, Lieb-Lindblad)

since 1995: explosion of quantum information theory, following the discovery of Shor's quantum algorithm for integer factoring  
- compressibility of "quantum information" (Schrödinger)  
- achievability of Holevo bound (HSW theorem), transitivity of quantum information (LSG theorem)  
- additivity questions: entanglement helps!  
- refinement of the notion along the lines of the classical case

Literature:

  \(\rightarrow\) available online


- M. Wilde, "From Classical to Quantum Shannon Theory" or arXiv:1106.1445 (available online)
1. Classical Data Compression

1.1 Probability distributions & convexity

- $X$: finite set, **alphabet** ($\in X$ are "symbols", "events", ...)
- $X$: random variable with range $\mathcal{X}$ and
  (probability) distribution $p:\mathcal{X} \to \mathbb{R}_+ = [0,\infty)$

  with normalization $\sum_{x \in \mathcal{X}} p(x) = 1$

  Also write $p(x) = p_x(x) = p_x$

**Example:** fair coin $\mathcal{X} = \{\text{heads}, \text{tails}\}$

  $p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$

- **Expectation value** $E[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$ (if $\mathcal{X} \subseteq \mathbb{R}$)

- **Joint distribution** $p_{xy} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$

  The values of the r.v. $XY$ are tuples $(x,y) \in \mathcal{X} \times \mathcal{Y}$

- **Marginal distributions** $p_x(x) = \sum_{y \in \mathcal{Y}} p_{xy}(x,y)$

- **Conditional distribution** $p_{x|y}(x|y) = \frac{p_{xy}(x,y)}{p_y(y)}$ for all $(x,y)$ with $p_y(y) > 0$

**Example:** $X \in \{0,1\}$ outcome of a fair coin toss

  $Y \in \{0,1,2\}$ sum of this coin toss plus a second fair coin toss

<table>
<thead>
<tr>
<th>$x \times y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

  $p_x(0) = \frac{1}{4}$, $p_x(1) = \frac{1}{2}$, $p_x(2) = \frac{1}{4}$

  $p_y(0) = \frac{1}{4}$, $p_y(1) = \frac{1}{2}$, $p_y(2) = \frac{1}{4}$

- r.v.s $X$ and $Y$ independent if $p_{xy}(x,y) = p_x(x) \cdot p_y(y) \ \forall x,y$

  (i.e. $p_{xy}(x|y) = p_x(x) \ \forall y$ with $p_y(y) > 0$)

  **Example:** $X$ and $(Y - X)$ from above are independent

---

Frequentist vs. Bayesian interpretation of probability:

*cf. e.g. [Mackay], inference*