Announcements

- Sign-up sheet
- Spots available in 11am recitations. Interested? E-mail Eduardo at sverdlin@mit.edu

Today

- 6.02 Recitations: Basics, administration, objectives
- Review of some basic probabilistic concepts
- Information
- Entropy
- Variable length binary coding – Huffman coding

Mujdat Cetin, 32-D616, mcetin@mit.edu, Office hours TBD
- Visiting professor on sabbatical here
- Worked at MIT from 2001-2005
- Taught 6.011 at MIT before

TA:
- 10-11 Max Dunitz?
- 11-12 Elaine Han

Recitations -- Tue/Thu 10-11 / 11-12
- will build upon some concepts covered in lectures
- will go through illustrative examples of concepts covered in lectures
- attend lectures, reviews (at least) the lecture slides before recitation
- please be interactive, ask and answer questions
- please come to my office hours for technical questions or any concerns you might have about this course

6.02 overview
- Intro to EECS through digital communications – very appropriate, because
digital communications contains many interesting problems on signals and systems and computer science
digital communications is ubiquitous these days!
- Pieces of the course: bits, signals, packets
  - Bits: information, coding
  - Signals: transmission in a physical channel
  - Packets: network communications
- Digital comms: Discrete set of messages rather than waveforms
  - remember the Morse alphabet example from lecture
  - Shannon: source emitting i.i.d. symbols

Start with a review of some basic probabilistic concepts
Probability Model (probabilistic experiment)

\[ U = \{ 1, 2, 3, 4, 5, 6 \} \]

- **Events**: subsets of \( U \).
  - Event \( A \) has occurred if the outcome of the experiment lies in the set \( A \).
  - Example: \( \text{The outcome is } \leq 3 \)
  - Example: \( \text{The outcome is odd} \)

- Probability measure:
  - Non-negative, sums to 1.

**Fair die**: \( p(a_i) = \frac{1}{6} \) \( \forall a_i \in U \)

---

Events form a algebra of sets

Consider events \( A, B \)

- **Union**: \( A \cup B \)/ \( A \text{ or } B \)/ \( A + B \)
- **Intersection**: \( A \cap B \)/ \( A \text{ and } B \)/ \( A - B \)

---

- Conditional probability \( P(A \mid B) \): prob. of event \( A \) given that \( B \) occurred.

\[ P(A \mid B) = \frac{P(AB)}{P(B)} \]  \( (*) \)

- Independence of two events: Whether one event has occurred or not does not change the probability of the other event.

\[ P(A \mid B) = P(A) \]

From \( (*) \) this is equivalent to \( P(AB) = P(A)P(B) \)
Intuition

- A: it rains  B: I get wet  - dependent
- A: it's Monday  B: I get wet  - independent

Example

**Bell of a 6-sided die**

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: outcome ≤ 2</td>
<td>( P(A) = \frac{1}{3} )</td>
</tr>
<tr>
<td>B: outcome odd</td>
<td>( P(B) = \frac{1}{2} )</td>
</tr>
<tr>
<td>( A \cap B )</td>
<td>( P(A \cap B) = \frac{1}{6} )</td>
</tr>
</tbody>
</table>

\[
P(A) = \frac{1}{3} \quad P(B) = \frac{1}{2} \quad P(A \cap B) = \frac{1}{6}
\]

\[
P(\text{A or B}) = P(A) + P(B) - P(\text{A and B})
\]

\[
P(\text{A or B}) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6}
\]

\[
P(\text{A or B}) = \frac{2}{3}
\]

Yes, independent.

What if A: outcome ≤ 2 \( \rightarrow \) not independent.

**Example**

**Bell of a 6-sided die**

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: outcome odd</td>
<td>( P(A) )</td>
</tr>
<tr>
<td>B: outcome &gt; 5</td>
<td>( P(B) )</td>
</tr>
</tbody>
</table>

Note: \( P(AB) = 0 \) - mutually exclusive events.

\[
P(A + B) = \frac{2}{3} = P(A) + P(B)
\]

In general, for any two events \( A, B \):

\[
P(A + B) = P(A) + P(B) - P(AB)
\]
Random variable: A random variable is a function on the outcome of a random experiment that assigns a number to each elementary outcome in the universe.

\[ X \subseteq \mathbb{R} \]

**Example:** Roll 2 dice: \( X \): Sum of two numbers showing up

Can write a probability mass fn.

\[
\begin{align*}
  P(x=x_j) &= p(x=x_j) \\
  p(x_j) &= \frac{1}{36} \quad \frac{2}{36} \quad \frac{1}{36} \\
  2 &\quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \\
\end{align*}
\]

The expected value or mean value of a r.v.

\[
E(x) = \sum_{j} x_j \cdot P(x=x_j)
\]

For the experiment above \( E(x) = 7 \)
Information

- Shannon: 1948 paper "A Mathematical Theory of Communication"
  1949 book "The Mathematical Theory of Communication"

- Information obtained by being told the outcome $S_i$ of a probabilistic experiment $S$:

$$I(S = S_i) = \log_2 \left( \frac{1}{p_S(S_i)} \right) \text{ bits}.$$ 

Measure of the uncertainty associated with this outcome prior to its being announced.

Less probable outcome/event happens $\Rightarrow$ more information revealed.

Information are degree of surprise.

Example:

Deck of 52 cards. - Drawing a card.

How much info do you get if I tell you it's a face card?

Uncertainty before I tell you anything:

$$\log_2 (52) = 5.708 \text{ bits}.$$ 

Uncertainty after I tell you it's a face card:

4 suits x 3 face cards/suit = 12 face cards

$$\log_2 (12) = 3.585 \text{ bits}.$$ 

Info I have provided:

$$\log_2 (52) - \log_2 (12) = \log_2 \left( \frac{52}{12} \right) = 2.115 \text{ bits}.$$ 

What if I tell you it's a face card & a spade?

Note: independent events

$$\frac{3}{52} = \frac{3}{12} \Rightarrow \text{ Info } = \log_2 \left( \frac{52}{12} \right) = 4.115 \text{ bits}.$$ 

Note: Info provided by occurrence of two independent events:

add info provided by each:

$$\log_2 \left( \frac{52}{12} \right) + \log_2 \left( \frac{4}{12} \right)$$
Entropy

The entropy $H(S)$ of random variable $S$ is the expected information obtained by learning the outcome of $S$.

$$H(S) = \sum_i p_S(s_i) \log_2 \left( \frac{1}{p_S(s_i)} \right) \text{ bits}$$

Measure of uncertainty of the random variable

- Maximal uncertainty when $p_S(s_i) = \frac{1}{N}$.
  $$\rightarrow H(S) = \log_2 N.$$

- Average amount of information that must be delivered to resolve the uncertainty of a random variable (source).

- Lower bound on the no. of binary digits needed to encode messages generated by a probabilistic source.

**Example:** Image coding - Consider images with 4 gray levels:

- $B$ - background
- $F_1, F_2, F_3$ - foreground pixel intensities

Simple binary coding:

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>00</td>
</tr>
<tr>
<td>$F_1$</td>
<td>01</td>
</tr>
<tr>
<td>$F_2$</td>
<td>10</td>
</tr>
<tr>
<td>$F_3$</td>
<td>11</td>
</tr>
</tbody>
</table>

We use 2 binary digits per symbol (pixel).

- Can we do better? Yes, if we exploit any knowledge about the source.

- "Most pixels in images are background pixels."
- Let's say we are told that with prob. 0.8 we have a background pixel, as different foreground intensities are equiprobable.
Intuition: use shorter codewords for more frequent symbols.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>F₁</td>
<td>10</td>
</tr>
<tr>
<td>F₂</td>
<td>110</td>
</tr>
<tr>
<td>F₃</td>
<td>111</td>
</tr>
</tbody>
</table>

Average codeword length:

$$0.8 \times 1 + \frac{0.2}{3} \times 2 + \frac{0.2}{3} \times 3 + \frac{0.2}{3} \times 3 = 1.333 \text{ binary digits}$$

So, we do better than simple binary coding!

Entropy provides lower bound on the no. of binary digits to be used on average to encode messages from a particular source. In this case:

$$H = 0.8 \log_2 \left( \frac{1}{0.8} \right) + 3 \cdot \frac{0.2}{3} \log_2 \left( \frac{3}{0.2} \right) = 1.0389 \text{ bits}$$

How can we design the minimum average length code for a particular source?