Noisy Channels

Inference

& Information measures for noisy channels
If we encode symbols from the ensemble

\[ \mathcal{A}_X = \{a, b, c, d\} \]
\[ \mathcal{P}_X = \{1/2, 1/4, 1/8, 1/8\} \]

using the symbol code

\[ \mathcal{C} = \{0, 10, 110, 111\}, \]

what is the probability \( p_1 \) that a bit plucked at random from the encoded stream is a 1?

A \( p_1 < \frac{1}{2} \)
B \( p_1 = \frac{1}{2} \)
C \( p_1 > \frac{1}{2} \)
\[ = \sum_{i} P_i \cdot f_i \]

\[ = \frac{1}{2} \times 0 + \frac{1}{4} \times \frac{1}{2} + \frac{1}{8} \times \frac{2}{3} + \frac{1}{8} \times 1 \]

\[ = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \times \frac{2}{3} \]

\[ = \frac{1}{8} \times \frac{8}{3} \]
\[ P_1 = \frac{\sum_i P_i \cdot N_i}{\sum_i P_i \cdot l_i} \]
\[ \frac{1}{2} \times 0 + \frac{1}{4} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 3 \]

\[ \frac{7}{4} \]

\[ \frac{3}{4} \]
\[\frac{1}{2} \times 0 + \frac{1}{4} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 3 = \frac{7}{14} = \frac{1}{2}\]
The ideal codelengths \( l_i^* \) are the information contents
\[
l_i^* = \log \frac{1}{p_i}
\]

The optimal symbol code’s expected length \( L \) satisfies
\[
H(X) \leq L < H(X) + 1
\]

Does that wrap up compression?

- Optimal symbol codes get within one bit per character of Shannon limit

Identical twins can help make compressors
Arithmetic coding
Arithmetic coding

achieves

\[ l(x) \leq \log_2 \frac{1}{P(x)} + 2 \]

where \( x \) is the whole file \( x_1 x_2 \ldots x_N \).
An example predictor

$$P(x_t | x_1 x_2 x_3 \ldots x_{t-1})$$

```
```

```
```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```
Data compression

An example predictor

\[ P(x_t \mid x_1x_2x_3...x_{t-1}) \]

..`I'm afraid I don't know one,' said Harriet, rather alarmed at the pro

PPM - 'prediction by partial match'
to her to be secure of a comfortable **provision**, 
and **proved** in the right, when so many people said 
was an irresistible **proof** of his great good sense, 
"Mrs. Bates, let me **propose** your venturing on one of 
was so great a personage in Highbury, that the **prospect**

Predictions proportional to **frequencies** (in this context)

**Six-gram model**
An example predictor

\[ P(x_t | x_1 x_2 x_3 \ldots x_{t-1}) \]

```
`I'm afraid I don't know one,' said Harriet, rather alarmed at the e_pro
  e_pro
  _pro
  pro
  ro
  o
```
Other uses for arithmetic coding

Efficient writing

Compression:

Text $\rightarrow$ Bit string
(preferably short)

Writing:

Text $\leftarrow$ Gesture
(preferably brief)
Dasher

www.inference.phy.cam.ac.uk/dasher/
Other uses for arithmetic coding

- Efficient writing

Compression:

Text $\rightarrow$ Bit string
(preferably short)

Writing:

Text $\leftarrow$ Gesture
(preferably brief)

- Efficient generation of random samples
$P_a = 0.01$

$P_b = 0.99$

$N$

$u \sim (0, 1)$

if $u < P_a \rightarrow a$

32 bits per real number
Arithmetic coding
\[ a \left\| \begin{array}{c} b \\ c \end{array} \right\| \quad P(x = ac) \]
AC method needs $H_2(0.01)$ bits per coin-toss.
THE MARSAGLIA RANDOM NUMBER CDROM
including the
DIEHARD BATTERY OF TESTS
OF RANDOMNESS

Research
Sponsored by
THE NATIONAL
SCIENCE
FOUNDATION
Grants
DMS-8807976
DMS-9206972

DEPARTMENT OF STATISTICS

and
SUPERCOMPUTER
COMPUTATIONS
RESEARCH
INSTITUTE

© 1995
George Marsaglia
All rights reserved

Florida State
UNIVERSITY

Professor
George Marsaglia
geo@stat.fsu.edu
The Marsaglia Random Number CD-ROM contains 5 billion random bits, divided into sixty 10-megabyte files.

It was developed and distributed under National Science Foundation Grants DMS-8807976 and DMS-9206972, at Florida State University.
Three cards
\[ P_a = 0.01 \]

\[ P_b = 0.99 \]

\[ N \]

\[ u \sim (0, 1) \]

if \( u < P_a \) \( \rightarrow a \)

\[ \frac{m}{2^{32}} \]
3 CARDS

\[ \begin{array}{ccc} 
  WW & BB & BW \\
\end{array} \]

\[ P \left( \text{other face is white} \mid \text{you see a white face uppermost} \right) = ? \]
A B C
p < p^2
p > p^2
\text{I don't know}
Always write down the probability of everything
Write down the probability of everything. © Steve Gull

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>Reverse</td>
<td>1/3</td>
<td>1/2</td>
</tr>
</tbody>
</table>

Then we can condition on data:

\[
P(\text{Reverse is } W \mid \text{Front is } W) = \frac{2}{3}
\]
What is the probability that the other face is white, given that you are now seeing a white face?
Communication over noisy channels

- inference
- information measures
**The rules**

- Gameshow host hides prize behind one door.
- Player chooses a door, e.g., door 1.
- Host opens another door, promising the prize will not be revealed.
- Player chooses to stick or switch, i.e., receive what's behind his original door, or the other closed door.
Player chooses door 1

Host opens door 3, revealing nothing, as promised

A  should stick
B  should switch
C  makes no difference
Z  don't know
Information measures

\[
\begin{align*}
H(X,Y) & \\
H(X) & \\
H(Y) & \\
H(X|Y) & I(X;Y) & H(Y|X)
\end{align*}
\]
Example Joint ensemble

\[ P(x,y) \]

<table>
<thead>
<tr>
<th>Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{32} )</td>
<td>( \frac{1}{32} )</td>
</tr>
<tr>
<td>2</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{32} )</td>
<td>( \frac{1}{32} )</td>
</tr>
<tr>
<td>3</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{16} )</td>
</tr>
<tr>
<td>4</td>
<td>( \frac{1}{4} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{32}$</td>
<td>$\frac{1}{32}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{32}$</td>
<td>$\frac{1}{32}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Marginal $P(y)$:

|   | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

$X$
"marginal entropy"

\[ H(X) = \frac{7}{4} \text{ bits} \]

\[ H(Y) = 2 \text{ bits} \]

\[ H(X,Y) = \sum_{(x,y)} P(x,y) \log_2 \frac{1}{P(x,y)} \]
\[ H(X) \]

\[ H(Y) \]

\[ H(X, Y) \]
\[ H(x) = \frac{7}{4} \text{ bits} \]
\[ H(y) = 2 \text{ bits} \]
\[ H(x,y) = \sum_{(x,y)} P(x,y) \log_2 \left( \frac{1}{P(x,y)} \right) = \frac{27}{8} \text{ bits} \]
Conditional probability:

\[ P(x \mid y) = \frac{P(x, y)}{P(y)} \]
$$P(x|y)$$

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/2</td>
<td>1/4</td>
<td>1/2</td>
<td>1/8</td>
</tr>
<tr>
<td>1/2</td>
<td>1/4</td>
<td>1/2</td>
<td>1/8</td>
<td>1/8</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
\[
\begin{align*}
H(X \mid y=1) &= \frac{7}{4} \\
H(X \mid y=2) &= \frac{7}{4} \\
\vdots & \ \\
\end{align*}
\]

= 2 \text{ bits}

= 0 \text{ bits}
\[ \begin{align*}
H(X|y=1) &= \frac{7}{4} \\
H(X|y=2) &= \frac{7}{4} \\
&= 2 \text{ bits} \\
&= 0 \text{ bits}
\end{align*} \]

\[ \text{Sim, } H(Y|X) = \frac{13}{8} \text{ bits} \]

\[ \text{Conditional Entropy} \]

\[ H(X|Y) = \sum_{y} p(y) H(X|y) \]

\[ = 1 \frac{1}{8} \text{ bits} \]
Conditional prob:

\[ P(x \mid y) = \frac{P(x,y)}{P(y)} \]

\[ H(X \mid Y) \leq H(X) \]
Mutual Information

\[ I(X;Y) = H(X) - H(X|Y) \] (Conditional)

\[ = H(Y) - H(Y|X) \]
Mutual information for the BSC

Consider \( f = 0.1 \).

Assume input distribution \( P_X = \{p_0, p_1\} = \{0.9, 0.1\} \).

Find

\[
I(X;Y) = H(Y) - H(Y | X) = H_2(0.18) - H_2(0.1) = 0.68 - 0.47 = 0.21 \text{ bits.}
\]
Inference for channels

Channel \( Q \) defines conditional probabilities

\[
P(y \mid x)
\]

If we choose an input distribution \( P(x) \), we have a joint distribution

\[
P(x, y) = P(x) \, P(y \mid x)
\]

with which we can do inference.
Channel $Q$ defines conditional probabilities

$$P(y \mid x)$$

If we choose an input distribution $P(x)$, we have a joint distribution

$$P(x, y) = P(x) P(y \mid x)$$

for which we can compute the mutual information

$$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$
Mutual information for the BSC

Consider $f = 0.1$.

Assume input distribution $\mathcal{P}_X = \{p_0, p_1\} = \{0.9, 0.1\}$.

Find
\[
I(X; Y) = H(Y) - H(Y \mid X)
= H_2(0.18) - H_2(0.1)
= 0.68 - 0.47
= 0.21 \text{ bits}.
\]
Recommended homework

Reading: Chapters 1-6; Advance reading: Chapters 8, 9, 10

See handout 2 on website for more recommended exercises

- Huffman programs huffman.p, huffman.py are on website
- also a 'bent coin' file 0010000.... as a compression benchmark