Information Theory

Master of Logic 2014/15
2nd Block Nov/Dec 2014

Some of these slides are copied from or heavily inspired by the University of Illinois at Chicago, ECE 534: Elements of Information Theory course given in Fall 2013 by Natasha Devroye
Thank you very much for the kind permission to re-use them here!
Christian Schaffner

- me
- pure mathematics at ETH Zurich
- PhD from Aarhus, Denmark
- research: quantum cryptography
- c.schaffner@uva.nl

- plays ultimate frisbee
Philip Schulz

- your teaching assistant
- PhD student @ILLC
- working on machine translation
- P.Schulz@uva.nl
Practicalities

- Final grade consists of 50-50:
- 7 weekly homework series, to be graded
- Final exam in week of 31/3/14 - 4/4/14: move to Tuesday, 16 December, but clashes with:
  * Lambda Calculus
  * Foundations of Neural and Cognitive Modelling
- Details on course homepage:
  http://homepages.cwi.nl/~schaffne/courses/inftheory/2014/
Expectations

We expect from you

- be on time
- code of honor (do not cheat)
- focus
- ask questions!

Why multitasking is bad for learning: https://medium.com/@cshirky/why-i-just-asked-my-students-to-put-their-laptops-away-7f5f7c50f368
Expectations

We expect from you
• be on time
• code of honor (do not cheat)
• focus
• ask questions!

You can expect from us
• be on time
• make clear what goals are
• listen to you and respond to email requests
• keep website up to date

Why multitasking is bad for learning: https://medium.com/@cshirky/why-i-just-asked-my-students-to-put-their-laptops-away-7f5f7c50f368
Questions ?
What is communication?
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.”

- C.E. Shannon, 1948

I want to send 1001

I think A sent 1001

A

B
Generic communication block diagram

What is communication?

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

-C.E. Shannon, 1948

I want to send 1001

I think A sent 1001

Generic communication block diagram

Source

Encoder

Channel

Decoder

Destination

Noise

Source coder

Channel coder

Channel
coder

Channel
decoder

Source decoder

Destination

Remove redundancy

Decode signals, detect/correct errors

Controlled adding of redundancy

Restore source

ECE 534 by Natasha Devroye
History of (wireless) communication

• Smoke signals
History of (wireless) communication

- **Smoke signals**
- **1861:** Maxwell’s equations

\[
\oint E \cdot dA = \frac{q_{\text{enc}}}{\varepsilon_0} \\
\oint B \cdot dA = 0 \\
\oint E \cdot ds = -\frac{d\Phi_B}{dt} \\
\oint B \cdot ds = \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt} + \mu_0 i_{\text{enc}}
\]
History of (wireless) communication

- **Smoke signals**
- **1861:** Maxwell’s equations
- **1900:** Guglielmo Marconi demonstrates wireless telegraph
History of (wireless) communication

- **Smoke signals**
- 1861: **Maxwell’s equations**
- 1900: **Marconi** demonstrates wireless telegraph
- 1920s: **Edwin Howard Armstrong** demonstrates FM radio
Big Open Questions

- mostly analog
- ad-hoc engineering, tailored to each application
- is there a general methodology for designing communication systems?
- can we communicate reliably in noise?
- how fast can we communicate?
Claude Elwood Shannon
1916 - 2001

• Father of Information Theory
• Graduate of MIT 1940:
  “An Algebra for Theoretical Genetics”
• 1941-1972: Scientist at Bell Labs
• 1958: Professor at MIT:
  When he returned to MIT in 1958, he continued to threaten corridor-walkers on his unicycle, sometimes augmenting the hazard by juggling. No one was ever sure whether these activities were part of some new breakthrough or whether he just found them amusing. He worked, for example, on a motorized pogo-stick, which he claimed would mean he could abandon the unicycle so feared by his colleagues ...

• juggling, unicycling, chess
• ultimate machine
History of (wireless) communication

- BITS!
- arguably, first to really define and use “bits”
- "He's one of the great men of the century. Without him, none of the things we know today would exist. The whole digital revolution started with him." -Neil Sloane, AT&T Fellow
• Introduced a new field: Information Theory

What is communication?

What is information?

How much can we compress information?

How fast can we communicate?
Main Contributions of Info Theory

Source coding

- source = random variable
- ultimate data compression limit is the source's entropy $H$
Main Contributions of Inf Theory

Source coding
- source = random variable
- ultimate data compression limit is the source’s entropy $H$

Channel coding
- channel = conditional distributions
- ultimate transmission rate is the channel capacity $C$
Main Contributions of Inf Theory

Source coding
• source = random variable
• ultimate data compression limit is the source’s entropy $H$

Channel coding
• channel = conditional distributions
• ultimate transmission rate is the channel capacity $C$

Reliable communication possible $\iff H < C$
Reactions to This Theory

- Engineers in disbelief
- Stuck in analogue world

How to approach the predicted limits?

Shannon says: can transmit at rates up to say 4Mbps over a certain channel without error. How to do it?
It Took 50 Years To Do It

How to approach the predicted limits?

review article by [Costello Forney 2006]
It Took 50 Years To Do It

- 50’s: algebraic codes

How to approach the predicted limits?

review article by [Costello Forney 2006]
It Took 50 Years To Do It

- 50’s: algebraic codes
- 60’s 70’s: convolutional codes

How to approach the predicted limits?

review article by [Costello Forney 2006]
It Took 50 Years To Do It

- 50’s: algebraic codes
- 60’s 70’s: convolutional codes
- 80’s: iterative codes (LDPC, turbo codes)

How to approach the predicted limits?

review article by [Costello Forney 2006]
It Took 50 Years To Do It

- 50’s: algebraic codes
- 60’s 70’s: convolutional codes
- 80’s: iterative codes (LDPC, turbo codes)
- 2009: polar codes

How to approach the predicted limits?

review article by [Costello Forney 2006]
It Took 50 Years To Do It

- 50’s: algebraic codes
- 60’s 70’s: convolutional codes
- 80’s: iterative codes (LDPC, turbo codes)
- 2009: polar codes

How to approach the predicted limits?

review article by [Costello Forney 2006]

Claude Shannon — Born on the planet Earth (Sol III) in the year 1916 A.D. Generally regarded as the father of the Information Age, he formulated the notion of channel capacity in 1948 A.D. Within several decades, mathematicians and engineers had devised practical ways to communicate reliably at data rates within 1% of the Shannon limit . . .

Encyclopedia Galactica, 166th ed.

Robert J. McEliece, Shannon Lecture 2004
Applications

- Communication Theory
- Computer Science (e.g. in cryptography)
- Physics (thermodynamics)
- Philosophy of Science (Occam’s Razor)
- Economics (investments)
- Biology (genetics, bio-informatics)
Topics Overview

• Entropy and Mutual Information
• Entropy Diagrams
• Perfectly Secure Encryption
• Data Compression
• Coding Theory
• Channel-Coding Theorem
• Zero-Error Information Theory
• Noisy-Channel Theorem
• Application to Machine Learning
Questions ?
### Example: Letter Frequencies

<table>
<thead>
<tr>
<th>$i$</th>
<th>$a_i$</th>
<th>$p_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>0.0575</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>0.0128</td>
</tr>
<tr>
<td>3</td>
<td>c</td>
<td>0.0263</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
<td>0.0285</td>
</tr>
<tr>
<td>5</td>
<td>e</td>
<td>0.0913</td>
</tr>
<tr>
<td>6</td>
<td>f</td>
<td>0.0173</td>
</tr>
<tr>
<td>7</td>
<td>g</td>
<td>0.0133</td>
</tr>
<tr>
<td>8</td>
<td>h</td>
<td>0.0313</td>
</tr>
<tr>
<td>9</td>
<td>i</td>
<td>0.0599</td>
</tr>
<tr>
<td>10</td>
<td>j</td>
<td>0.0006</td>
</tr>
<tr>
<td>11</td>
<td>k</td>
<td>0.0084</td>
</tr>
<tr>
<td>12</td>
<td>l</td>
<td>0.0335</td>
</tr>
<tr>
<td>13</td>
<td>m</td>
<td>0.0235</td>
</tr>
<tr>
<td>14</td>
<td>n</td>
<td>0.0596</td>
</tr>
<tr>
<td>15</td>
<td>o</td>
<td>0.0689</td>
</tr>
<tr>
<td>16</td>
<td>p</td>
<td>0.0192</td>
</tr>
<tr>
<td>17</td>
<td>q</td>
<td>0.0008</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
<td>0.0508</td>
</tr>
<tr>
<td>19</td>
<td>s</td>
<td>0.0567</td>
</tr>
<tr>
<td>20</td>
<td>t</td>
<td>0.0706</td>
</tr>
<tr>
<td>21</td>
<td>u</td>
<td>0.0334</td>
</tr>
<tr>
<td>22</td>
<td>v</td>
<td>0.0069</td>
</tr>
<tr>
<td>23</td>
<td>w</td>
<td>0.0119</td>
</tr>
<tr>
<td>24</td>
<td>x</td>
<td>0.0073</td>
</tr>
<tr>
<td>25</td>
<td>y</td>
<td>0.0164</td>
</tr>
<tr>
<td>26</td>
<td>z</td>
<td>0.0007</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>0.1928</td>
</tr>
</tbody>
</table>

Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the area of white squares.
Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from The Frequently Asked Questions Manual for Linux). The picture shows the probabilities by the areas of white squares.

Figure 2.2. The probability distribution over the 27×27 possible bigrams $xy$ in an English language document, The Frequently Asked Questions Manual for Linux.
Example: Surprisal Values

from http://www.umsl.edu/~fraundorfp/egsurpri.html

<table>
<thead>
<tr>
<th>situation</th>
<th>probability p = 1/2^#bits</th>
<th>surprisal #bits = ln₂[1/p]</th>
</tr>
</thead>
<tbody>
<tr>
<td>one equals one</td>
<td>1</td>
<td>0 bits</td>
</tr>
<tr>
<td>wrong guess on a 4-choice question</td>
<td>3/4</td>
<td>ln₂[3/4] ~ 0.415 bits</td>
</tr>
<tr>
<td>correct guess on true-false question</td>
<td>1/2</td>
<td>ln₂[2] = 1 bit</td>
</tr>
<tr>
<td>correct guess on a 4-choice question</td>
<td>1/4</td>
<td>ln₂[4] = 2 bits</td>
</tr>
<tr>
<td>seven on a pair of dice</td>
<td>6/6² = 1/6</td>
<td>ln₂[6] ~ 2.58 bits</td>
</tr>
<tr>
<td>snake-eyes on a pair of dice</td>
<td>1/6² = 1/36</td>
<td>ln₂[36] ~ 5.17 bits</td>
</tr>
<tr>
<td>random character from the 8-bit ASCII set</td>
<td>1/256</td>
<td>ln₂[2⁸] = 8 bits = 1 byte</td>
</tr>
<tr>
<td>N heads on a toss of N coins</td>
<td>1/2ᴺ</td>
<td>ln₂[2ᴺ] = N bits</td>
</tr>
<tr>
<td>harm from a smallpox vaccination</td>
<td>~1/1,000,000</td>
<td>~ln₂[10⁶] ~ 19.9 bits</td>
</tr>
<tr>
<td>win the UK Jackpot lottery</td>
<td>1/13,983,816</td>
<td>~23.6 bits</td>
</tr>
<tr>
<td>RGB monitor choice of one pixel's color</td>
<td>1/2⁶⁵³ ~ 5.9×10⁻⁸</td>
<td>ln₂[2⁶⁵³] = 24 bits</td>
</tr>
<tr>
<td>gamma ray burst mass extinction event TODAY!</td>
<td>&lt;1/(10⁹×365) ~ 2.7×10⁻¹²</td>
<td>hopefully &gt;38 bits</td>
</tr>
<tr>
<td>availability to reset 1 gigabyte of random access memory</td>
<td>1/2^{2⁸E⁹} ~ 10⁻².⁴E⁹</td>
<td>8×10⁹ bits ~ 7.6×10⁻¹⁴ J/K</td>
</tr>
<tr>
<td>choices for 6×10²³ Argon atoms in a 24.2L box at 295K</td>
<td>~1/2⁻¹.⁶¹E²⁵ ~ 10⁻⁴.⁸E²⁴</td>
<td>~1.⁶¹×10²⁵ bits ~ 155 J/K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>a_i</th>
<th>p_i</th>
<th>h(p_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>.0575</td>
<td>4.1</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>.0128</td>
<td>6.3</td>
</tr>
<tr>
<td>3</td>
<td>c</td>
<td>.0263</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
<td>.0285</td>
<td>5.1</td>
</tr>
<tr>
<td>5</td>
<td>e</td>
<td>.0913</td>
<td>3.5</td>
</tr>
<tr>
<td>6</td>
<td>f</td>
<td>.0173</td>
<td>5.9</td>
</tr>
<tr>
<td>7</td>
<td>g</td>
<td>.0133</td>
<td>6.2</td>
</tr>
<tr>
<td>8</td>
<td>h</td>
<td>.0313</td>
<td>5.0</td>
</tr>
<tr>
<td>9</td>
<td>i</td>
<td>.0599</td>
<td>4.1</td>
</tr>
<tr>
<td>10</td>
<td>j</td>
<td>.0006</td>
<td>10.7</td>
</tr>
<tr>
<td>11</td>
<td>k</td>
<td>.0084</td>
<td>6.9</td>
</tr>
<tr>
<td>12</td>
<td>l</td>
<td>.0335</td>
<td>4.9</td>
</tr>
<tr>
<td>13</td>
<td>m</td>
<td>.0235</td>
<td>5.4</td>
</tr>
<tr>
<td>14</td>
<td>n</td>
<td>.0596</td>
<td>4.1</td>
</tr>
<tr>
<td>15</td>
<td>o</td>
<td>.0689</td>
<td>3.9</td>
</tr>
<tr>
<td>16</td>
<td>p</td>
<td>.0192</td>
<td>5.7</td>
</tr>
<tr>
<td>17</td>
<td>q</td>
<td>.0008</td>
<td>10.3</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
<td>.0508</td>
<td>4.3</td>
</tr>
<tr>
<td>19</td>
<td>s</td>
<td>.0567</td>
<td>4.1</td>
</tr>
<tr>
<td>20</td>
<td>t</td>
<td>.0706</td>
<td>3.8</td>
</tr>
<tr>
<td>21</td>
<td>u</td>
<td>.0334</td>
<td>4.9</td>
</tr>
<tr>
<td>22</td>
<td>v</td>
<td>.0069</td>
<td>7.2</td>
</tr>
<tr>
<td>23</td>
<td>w</td>
<td>.0119</td>
<td>6.4</td>
</tr>
<tr>
<td>24</td>
<td>x</td>
<td>.0073</td>
<td>7.1</td>
</tr>
<tr>
<td>25</td>
<td>y</td>
<td>.0164</td>
<td>5.9</td>
</tr>
<tr>
<td>26</td>
<td>z</td>
<td>.0007</td>
<td>10.4</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>.1928</td>
<td>2.4</td>
</tr>
</tbody>
</table>

\[ \sum_i p_i \log_2 \frac{1}{p_i} = 4.1 \]

Table 2.9. Shannon information contents of the outcomes a–z.