ESE 523
Information Theory

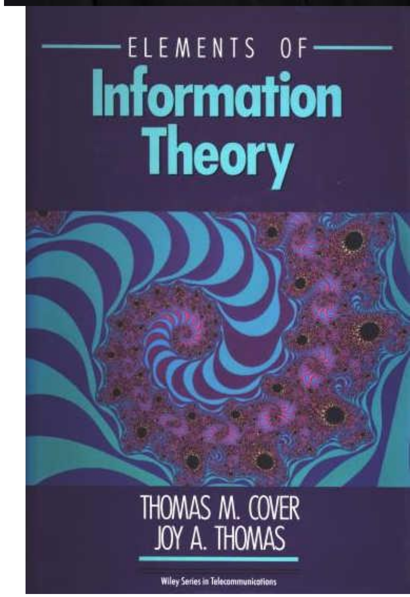
Joseph A. O'Sullivan
Samuel C. Sachs Professor
Electrical and Systems Engineering
Washington University
211 Urbauer Hall
2120E Green Hall
314-935-4173 (Lynda Markham Answers)
jao@wustl.edu

Outline

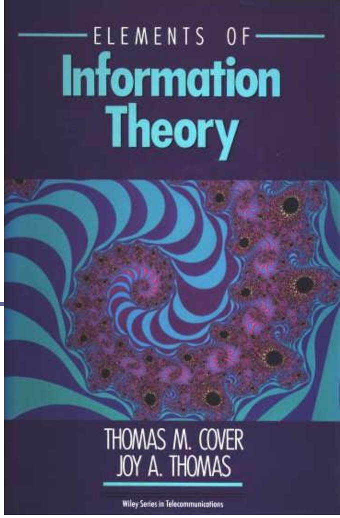
- Names, introduction
- Course Description
- Tentative Schedule
- Topics

Course Description

- Textbook: Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Second Edition, New York: Wiley and Sons, 2006.
- Time and place: TuTh 8:30-10:00 a.m., Green Hall 0159
- Office hours: TBD
- Catalog Description
  - Discrete source and channel model, definition of information rate and channel capacity, coding theorems for sources and channels, encoding and decoding of data for transmission over noisy channels. Corequisite: ESE 520 or equivalent.
- Problem Set Solutions
  - Solutions will be made available. Use of back files of any kind in the solution of problem sets is strictly forbidden. See Course Policy Statement.



Tentative Schedule



- Chapter 1 and preview, Lecture 1
- Chapter 2: Entropy, relative entropy, and mutual information 3 lectures
- Chapter 3: Asymptotic equipartition property, 1 lecture
- Chapter 4: Entropy rates of a stochastic process, 2 lectures
- Chapter 5: Data compression, 3 lectures
- Chapter 6: Gambling and data compression, 2 lectures
- Chapter 7: Channel capacity, 3 lectures
- Chapter 8: Differential entropy, 1 lecture
- Chapter 12: Maximum entropy, 1 lecture
- Chapter 9: Gaussian channel capacity, 2 lectures
- Chapter 10: Rate-distortion theory, 2 lectures
- Chapter 11: Information theory and statistics, 4 lectures
- Chapter 13: Universal source coding, 2 lectures
- Chapter 15: Network information theory, 1 lecture
- Total: 28 lectures?!

Schedule Comments

- Mostly we work straight through the book. We skip some sections.
- Chapter 6: Gambling is fun
- Chapter 14: Kolmogorov complexity is hard
- Chapter 15: Network information theory continues to be a hot research area

J. L. Kelly, Jr.;
See also Ed Thorp

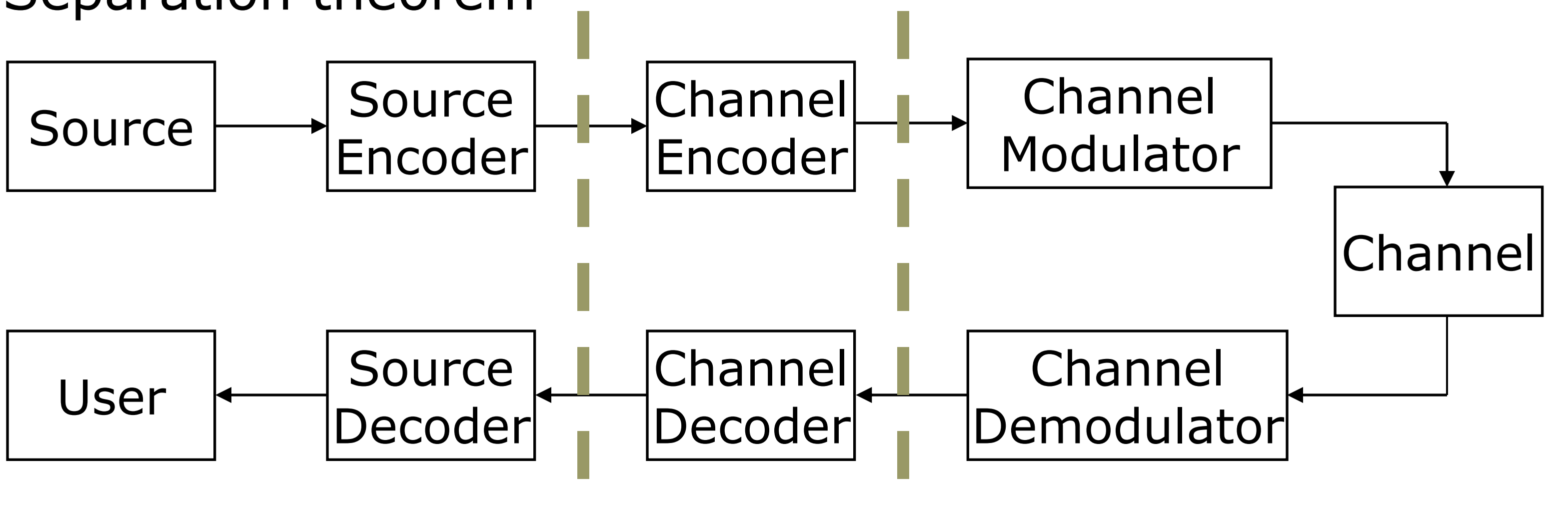
Andrey Kolmogorov

What is information theory?

- Information Theory contributes to and draws results from a number of other fields including:
  - Communication Theory
  - Probability Theory
  - Statistics: Estimation and Detection Theory
  - Physics
  - Computer Science
  - Economics

Origins in Communication Theory

- Shannon 1948. The Mathematical Theory of Communication. BSTJ.
- Channel capacity: fundamental bound on bits per second; channel coding deals with approaching this rate
- Source encoding: bound on number of bits to compress a source
- Rate-distortion theory: bound on number of bits to compress a source subject to a distortion constraint
- Separation theorem



Complexity Theory

- Time-complexity
- Space-complexity
- Kolmogorov complexity
  - More closely related to information theory and equals the length of the smallest program that can be written to reproduce the data.
  - This is not computable.
  - For random data, this quantity is approximately equal to the entropy.

Physics

- In statistical physics (statistical mechanics), the fundamental quantities studied are energy and entropy
- Second Law: Entropy of systems must increase (subject to other constraints)
- Quantum Information Theory
- Information theory and black holes: apparently information is preserved in the event horizon

Statistics

- The Fisher information determines a lower bound on the variance of any estimator.
- Relative entropy determines the exponent in the probability of error in optimal detection algorithms as the data set grows.
- Large deviations quantifies the probability of rare events.

Combinatorics

- Given a discrete alphabet with m elements
- n independent and identically distributed (i.i.d.) elements are drawn (sampling with replacement)
- How many ways are there to get k_1, k_2, … k_m of the elements ($k_1+k_2+ \ldots +k_m=n$)?
- Call this a *type*. What is the probability of getting one outcome in such a type?
- What is the probability of getting a type?
- What type is most likely?
- What is the probability of getting a type far from the most likely?

Combinatorics: Answers part 1

- How many ways are there to get $k_1, k_2, \ldots k_m$ of the elements ($k_1+k_2+ \ldots +k_m=n$)?
- What is the probability of getting one outcome in such a type?
- Lemma:

$$\frac{1}{k}\ln k! \underset{k\to\infty}{\rightarrow} \ln k - 1$$

$$\binom{n}{k_1 k_2 \ldots k_m} = \frac{n!}{k_1!k_2!\ldots k_m!}$$

$$\ln k! = \sum_{i=1}^{k} \ln i$$

$$\int_1^k (\ln x)\,dx < \sum_{i=1}^{k} \ln i < \int_1^{k+1} (\ln x)\,dx$$

$$k\ln k - k + 1 < \sum_{i=1}^{k} \ln i < (k+1)\ln(k+1) - (k+1) + 1$$

$$\sum_{i=1}^{k} \ln i = k\left(\ln k - 1 + o(k)\right)$$

Combinatorics: Answers part 1

- How many ways are there to get $k_1, k_2, \dots k_m$ of the elements $(k_1+k_2+\dots+k_m=n)$?
- What is the probability of getting one outcome in such a type?

$$\binom{n}{k_1 k_2 \dots k_m} = \frac{n!}{k_1! k_2! \dots k_m!}$$

$$= 2^{n\left(-\frac{k_1}{n}\log\frac{k_1}{n} - \frac{k_2}{n}\log\frac{k_2}{n} \dots - \frac{k_m}{n}\log\frac{k_m}{n} + o(n)\right)}$$

$$= 2^{nh\left(\frac{k_1}{n}, \frac{k_2}{n}, \dots, \frac{k_m}{n}\right) + no(n)}$$

$$h(p_1, p_2, \dots, p_m) = \sum_{i=1}^{m} -p_i \log p_i$$

$$q_1^{k_1} q_2^{k_2} \dots q_m^{k_m} = 2^{n(p_1 \log q_1 + p_2 \log q_2 \dots + p_m \log q_m)}$$

$$= 2^{-n\left(p_1 \log\frac{p_1}{q_1} + p_2 \log\frac{p_2}{q_2} \dots + p_m \log\frac{p_m}{q_m} + h(p_1, p_2, \dots, p_m)\right)}$$

$$p_1 = \frac{k_1}{n} \quad p_2 = \frac{k_2}{n} \quad \dots \quad p_m = \frac{k_m}{n}$$

- *Entropy of empirical distribution determines the number of elements in a type (exponent is the entropy)*
- *Probability of one outcome is determined by the type and the probability of each outcome*

Combinatorics : Answers part 2

- What is the probability of getting a type? Answer: determined by relative entropy between type and probability
- What type is most likely? Answer: $\mathbf{p}=\mathbf{q}$
- What is the probability of getting a type far from the most likely? Answer: determined by the relative entropy to the closest type in the unlikely set

$$q_1^{k_1} q_2^{k_2} \ldots q_m^{k_m} \binom{n}{k_1 k_2 \ldots k_m}$$

$$= 2^{-n\left(p_1 \log\frac{p_1}{q_1} + p_2 \log\frac{p_2}{q_2} \ldots + p_m \log\frac{p_m}{q_m} + o(n)\right)}$$

$$= 2^{-n\mathrm{D}(\mathbf{p}\|\mathbf{q}) + no(n)}$$

$$p_1 = \frac{k_1}{n} \quad p_2 = \frac{k_2}{n} \quad \ldots \quad p_m = \frac{k_m}{n}$$

$$2^{-n\min_{\mathbf{p}\in\Pi_0} \mathrm{D}(\mathbf{p}\|\mathbf{q}) + no(n)}$$

- *Relative entropy between empirical distribution and probability distribution determines probability of a type (exponential in relative entropy)*

Combinatorics : Experiments

- Bernoulli p (binary, $P(X_i=1)=p$
- 50 i.i.d. draws
- Repeat 1,000,000 times
- See Matlab Code

Matlab Code

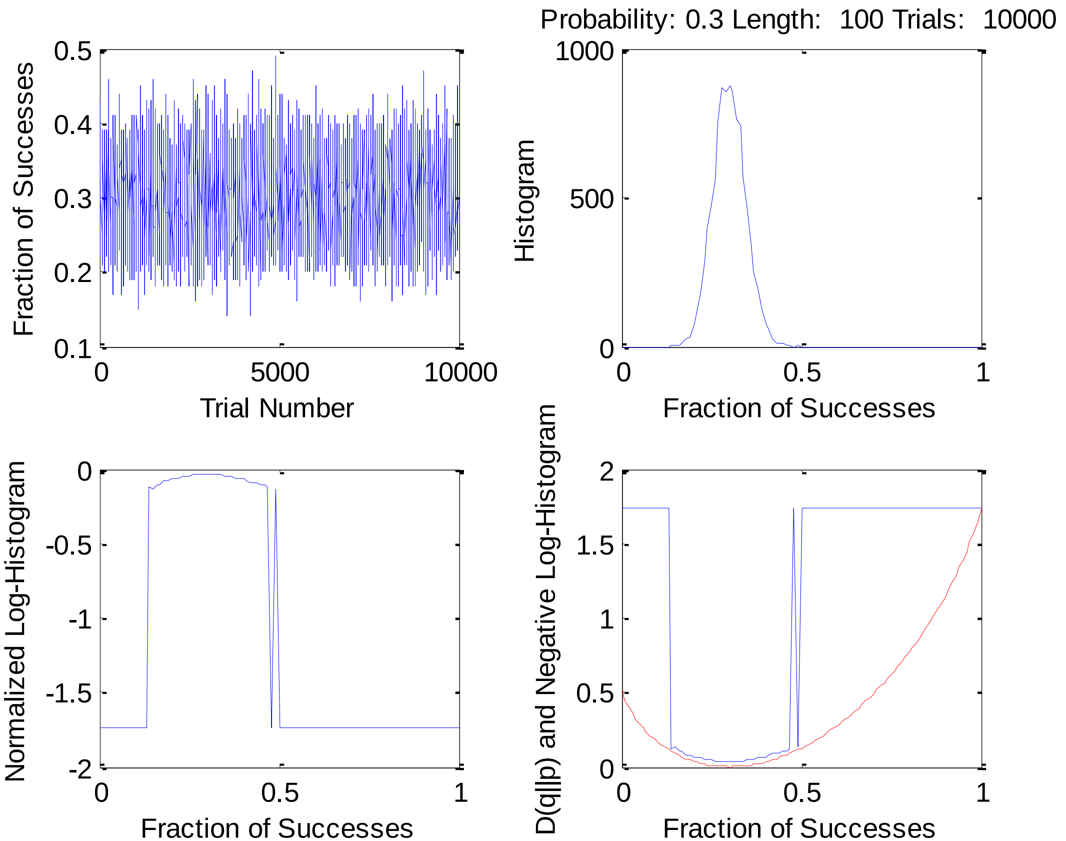
```
function [histout]=binhist(p,n,m)
%  histout=binhist(p,n,m)
%  p=Bernoulli probability
%  n=length of vector
%  m=number of trials
%  Program runs m trials and compares the normalized log histogram
%  to the relative entropy that bounds the probability of large devaiations.
%
%   This file was created by Joseph A. O'Sullivan
%   August 28, 2003
%   copyright 2003
x=rand(n,m);
y=(sign(x-1+p)+1)/2;
ky=sum(y,1)/n;
q=0:1/n:1;
histout=hist(ky,q);
figure
subplot(2,2,1)
plot([1:m],ky)
xlabel('Trial Number')
ylabel('Fraction of Successes')
```

continued …

Matlab Code

```
subplot(2,2,2)
plot(q,histout)
xlabel('Fraction of Successes')
ylabel('Histogram')
titlestring=['Probability: ', num2str(p), ' Length:  ', num2str(n), ' Trials:  ', num2str(m)];
title(titlestring)
subplot(2,2,3)
loghist=max(min(log2(p),log2(1-p)),log2(histout/m)/n);
plot(q,loghist)
xlabel('Fraction of Successes')
ylabel('Normalized Log-Histogram')
subplot(2,2,4)
dpq=q(2:n).*log2(q(2:n)/p)+(1-q(2:n)).*log2((1-q(2:n))/(1-p));
dpq=[-log2(1-p) dpq];
dpq=[dpq -log2(p)];
plot(q,dpq,'r')
hold on
plot(q,-loghist)
xlabel('Fraction of Successes')
ylabel('D(q||p) and Negative Log-Histogram')
```

Example Result



- Histogram of fraction of successes looks reasonable
- Probability of large deviations is exponentially small