

Noisy Channels

Inference

& Information measures
for noisy channels

Q: If we encode symbols from the ensemble

$$\mathcal{A}_X = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$$

$$\mathcal{P}_X = \{1/2, 1/4, 1/8, 1/8\}$$

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 $p_1 > \frac{1}{2}$

h_i	P_i		
1	$\frac{1}{2}$	a	○
2	$\frac{1}{4}$	b	○
3	$\frac{1}{8}$	c	○
4	$\frac{1}{8}$	d	

f_i	l_i
0	1
$\frac{1}{2}$	2
$\frac{2}{3}$	3
1	3

A	4
B	1
C	1
Z	

$$= \sum_i p_i f_i$$

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

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \times \frac{2}{3}$$

$$= \frac{1}{8} \times \frac{8}{3}$$

$$P_1 = \frac{\sum_i p_i n_i}{\sum_i p_i h_i}$$

$$= \frac{1}{2} \times 0 + \frac{1}{4} \times 1 + \overbrace{\frac{1}{8} \times 2}^{\frac{1}{4}} + \overbrace{\frac{1}{8} \times 3}^{\frac{3}{8}}$$

$$\frac{7}{4}$$

$$\frac{1}{2} \times 0 + \frac{1}{4} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 3 = \frac{7}{8} = \frac{1}{2}$$

$$\frac{7}{4}$$

The image shows a chalkboard with a mathematical calculation. The top part of the board contains the equation $\frac{1}{2} \times 0 + \frac{1}{4} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 3 = \frac{7}{8} = \frac{1}{2}$. Above the terms $\frac{1}{8} \times 2$ and $\frac{1}{8} \times 3$, there are curly braces with $\frac{1}{4}$ and $\frac{3}{8}$ written above them, respectively. Below the equation, a horizontal line is drawn. Underneath this line, the fraction $\frac{7}{4}$ is written. To the right of the main equation, the fraction $\frac{7}{8}$ is written above a horizontal line, with $\frac{7}{4}$ written below it. To the right of this, the fraction $\frac{1}{2}$ is written and underlined twice.

Last time - symbol codes

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

uniform bounded at_mouse at_cross

Colors: 28 Truncate: 6 At: < > 6 Slide: < > -2

Reset magnifn: 0 width: 1.0 centre: 0.5 0.0 1.0

english1 english2 jack Alphabet 3 adaptive 0.4 0.4 0.1

Alphabet 2 adaptive 0.5 0.5

Active Pushiness 1 ps Clean VariableL DelDisPel Reinstate Active Pushiness 1 ps

:ba 2 :01 2

Quit Hide Show OneCanvas Dump Mousepad Help

Arithmetic coding

achieves

$$l(\mathbf{x}) \leq \log_2 \frac{1}{P(\mathbf{x})} + 2$$

where \mathbf{x} is the whole file $x_1 x_2 \dots x_N$.

Data compression

● An example predictor

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

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

Data compression

● An example predictor

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

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

PPM - 'prediction by partial match'

to her to be secure of a comfortable_provision,
and be_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

Data compression

● An example predictor

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

`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



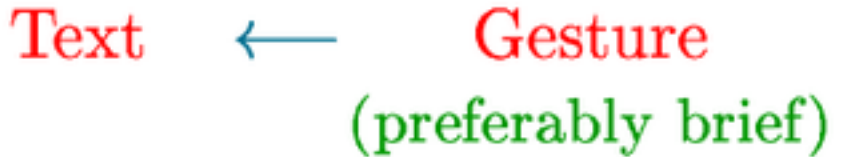
Dasher

Information-Efficient Text Entry

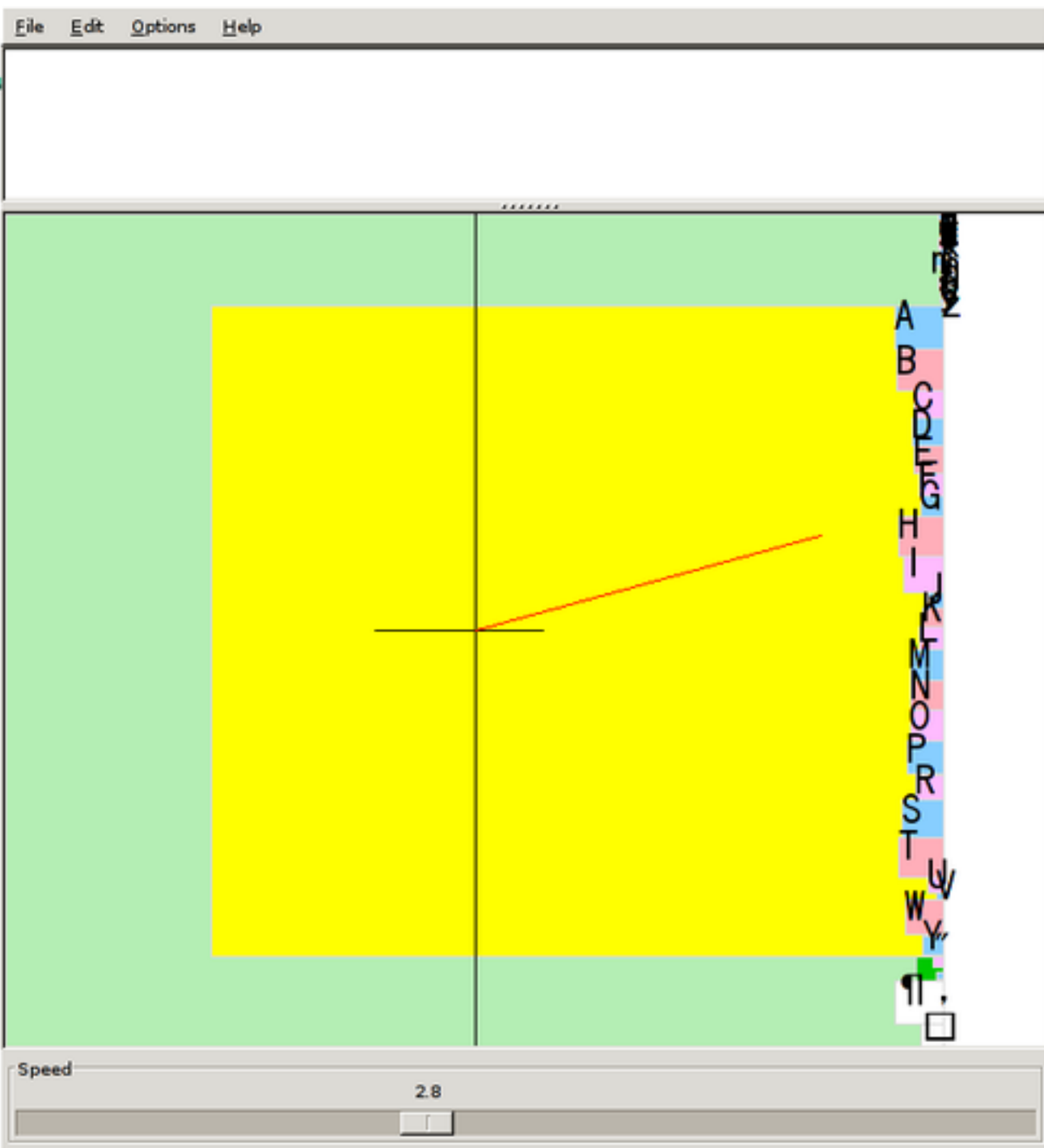
Compression:



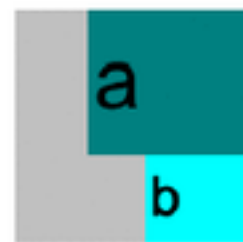
Writing:



Dasher



Other uses for arithmetic coding



Dasher

Information-Efficient Text Entry

● Efficient writing

Compression:

Text \longrightarrow Bit string
(preferably short)

Writing:

Text \longleftarrow 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

uniform bounded at_mouse at_cross Colors: 28 Truncate: 6 At: < > 6 Slide: < > -2

Reset magnifn: 0 width: 1.0 centre: 0.5 0.0 1.0

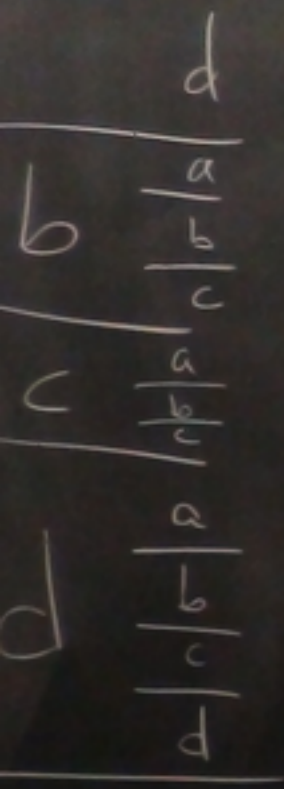
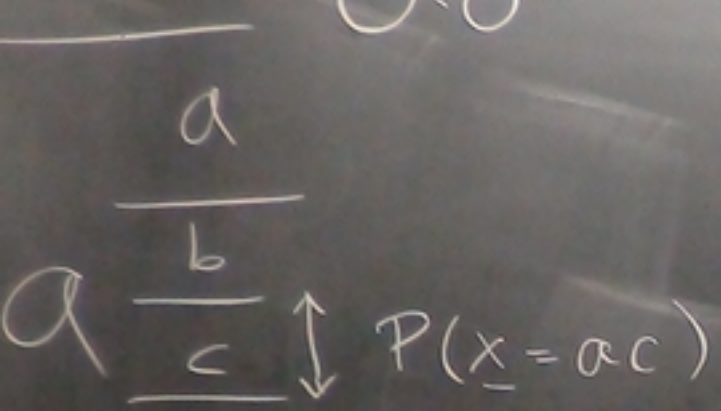
english1 english2 jack Alphabet 3 adaptive 0.4 0.4 0.1 Alphabet 2 adaptive 0.5 0.5

Active Pushiness 1 ps Clean VariableL DelDisPel Reinstate Active Pushiness 1 ps

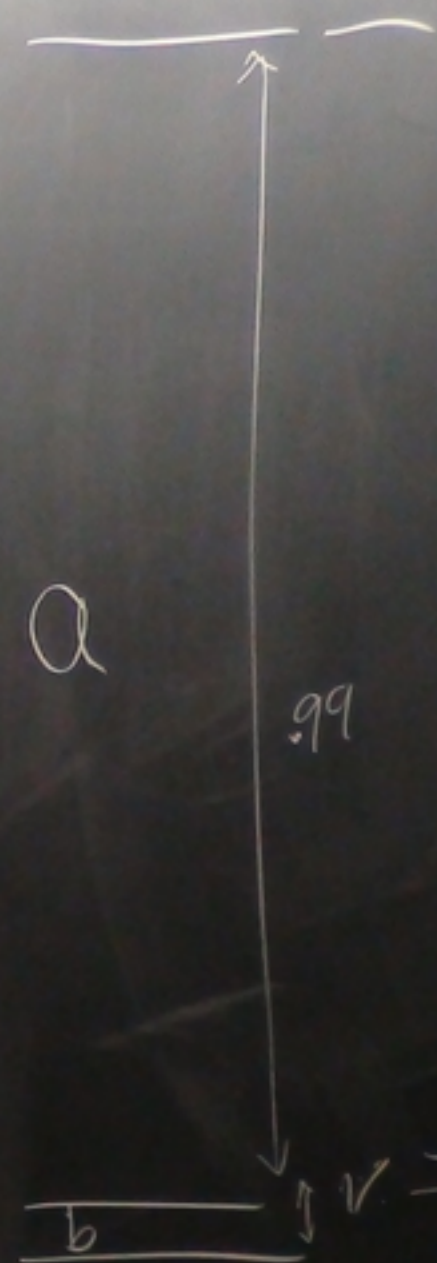
:ba 2 :01 2

Quit Hide Show OneCanvas Dump Mousepad Help

0.0



1.0



ant
ab

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

Professor
George Marsaglia
geo@stat.fsu.edu


Florida State
UNIVERSITY

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

© 1995
George Marsaglia
All rights reserved



DEPARTMENT
OF STATISTICS

and

SUPERCOMPUTER
COMPUTATIONS
RESEARCH
INSTITUTE

Professor
George Marsaglia
geo@stat.fsu.edu

A screenshot of a web browser window. The address bar shows the URL "http://stat.fsu.edu/pub/diehard/cdrom/". The browser has several tabs open, including "TinyURL!", "timesTrains", "google", and "go to referrer". The current page is titled "Index of /pub/diehard/cdr...". The main content is a directory listing with columns for "Name", "Last modified", "Size", and "Description". The listing includes a "Parent Directory" link and a series of files named "bits.01" through "bits.09", each with a date of "09-Dec-1995", a time, and a size of "9.5M".

Name	Last modified	Size	Description
Parent Directory		-	
bits.01	09-Dec-1995 10:21	9.5M	
bits.02	09-Dec-1995 10:23	9.5M	
bits.03	09-Dec-1995 10:24	9.5M	
bits.04	09-Dec-1995 10:26	9.5M	
bits.05	09-Dec-1995 10:27	9.5M	
bits.06	09-Dec-1995 10:29	9.5M	
bits.07	09-Dec-1995 10:30	9.5M	
bits.08	09-Dec-1995 10:32	9.5M	
bits.09	09-Dec-1995 10:33	9.5M	

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

white - 1f

brown - 2f

red - 3f

$$P_a = 0.01$$

$$P_b = 0.99$$

$$\frac{m}{2^{32}}$$

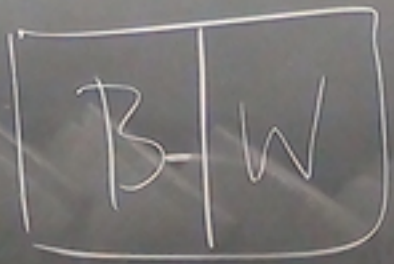
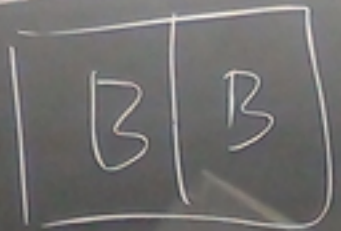
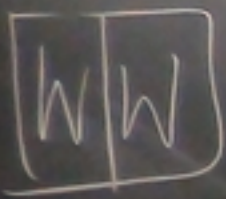
N

$$u \sim (0, 1)$$

32 bits per real number

$$\text{if } u < P_a \rightarrow a$$

3 CARDS



$$P(\text{other face is white} \mid \text{you see a white face upmost}) = ?$$

A $p < \frac{1}{2}$ 0

B $p > \frac{1}{2}$ 7

C $p = \frac{1}{2}$ 6

Z don't know 1

Always write down the probability of everything

write down the probability of everything

© Steve Gull

reverse

	front W	B
W	$\frac{1}{3}$	$\frac{1}{6}$
B	$\frac{1}{6}$	$\frac{1}{3}$

then we can condition on data

$$P(\text{reverse is W} \mid \text{front} = \text{W}) = \underline{\underline{\frac{2}{3}}}$$

Three cards

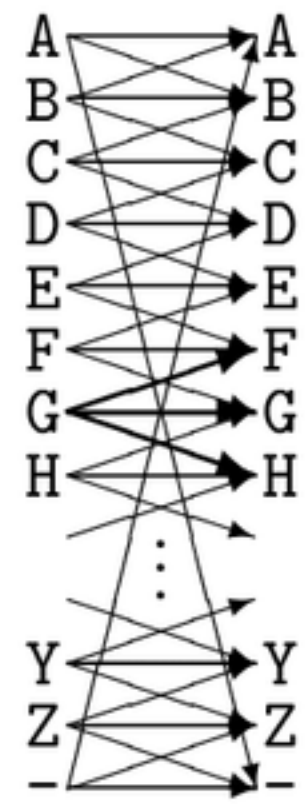
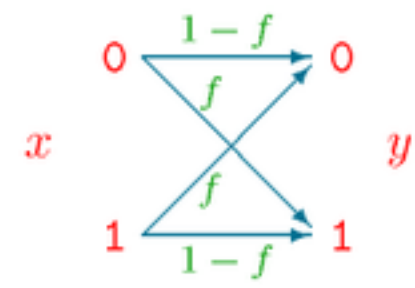
Q

What is the probability that
the other face is white,
given that you are now seeing a white face?

Communication over noisy channels

Noisy channels

- inference
- information measures



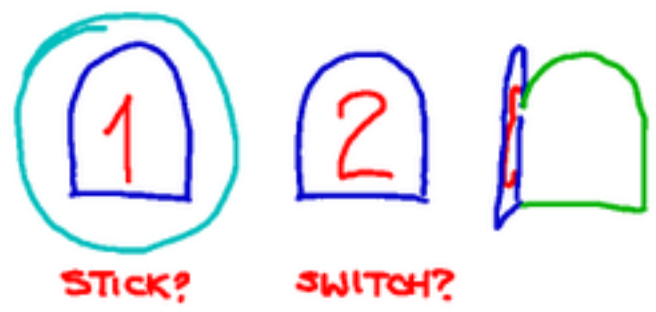
THREE DOORS



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

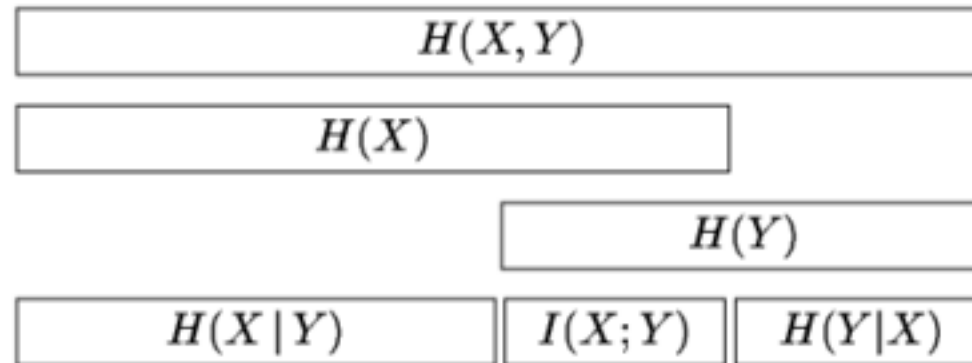
Q

- 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



Example

Marginal

$P(y)$

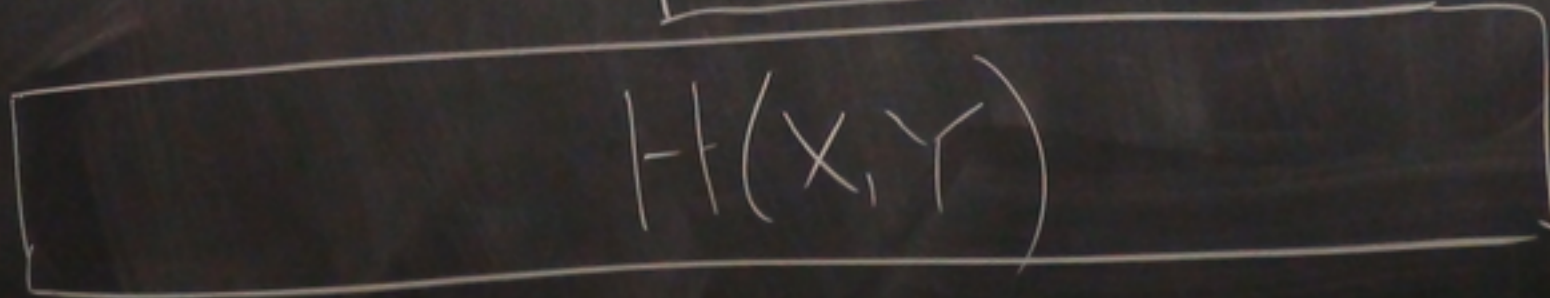
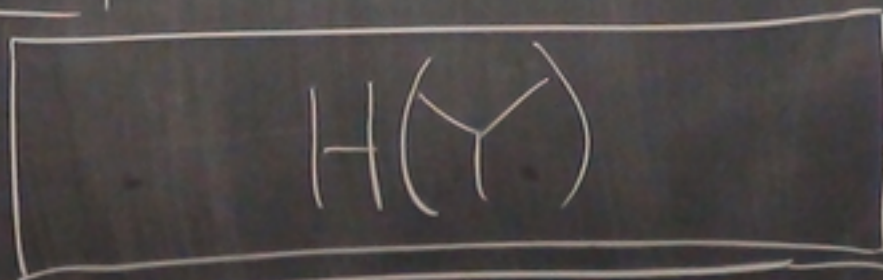
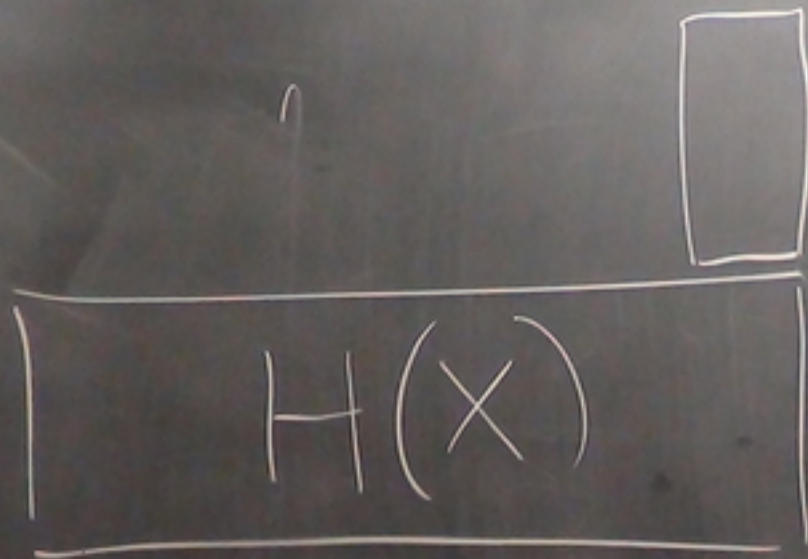
	X				
	1	2	3	4	
1	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
4	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$
		$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	

"marginal
entropy"

$$H(X) = 7/4 \text{ bits}$$

$$H(Y) = 2 \text{ bits}$$

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



$$H(X) = 7/4 \text{ bits}$$

$$H(Y) = 2 \text{ bits}$$

$$H(X, Y) = \sum_{(x, y)} P(x, y) \log_2 \frac{1}{P(x, y)} = \frac{27}{8} \text{ bits}$$

Conditional probz

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

$$H(\text{X} | y=1) = 7/4$$

$$H(\text{X} | y=2) = 7/4$$

$$\vdots = 2 \text{ bits}$$

$$\vdots = 0 \text{ bits}$$



No smoking

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

CONDITIONAL ENTROPY

$$\begin{aligned} H(X|y=1) &= \frac{7}{4} \\ H(X|y=2) &= \frac{7}{4} \\ &= 2 \text{ bits} \\ &\vdots \\ &= 0 \text{ bits} \end{aligned}$$

$$H(X|Y)$$

$$\begin{aligned} &= \sum_y P(y) H(X|y) \\ &= \frac{11}{8} \text{ bits} \end{aligned}$$

Conditional prob^y

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

$$H(X|Y) \leq H(X)$$

MUTUAL INFORMATION



$$I(X; Y) = H(X) - H(X|Y) \quad \left(\text{Condition} \right)$$

$$H(X)$$

$$H(Y|X)$$

$$= H(Y) - H(Y|X) \quad \left(P(X) \right)$$

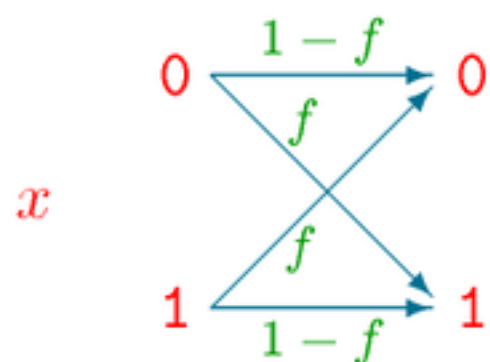
$$H(X|Y)$$

$$H(Y)$$

$$H(X, Y)$$

$$H(X|Y)$$

Mutual information for the BSC



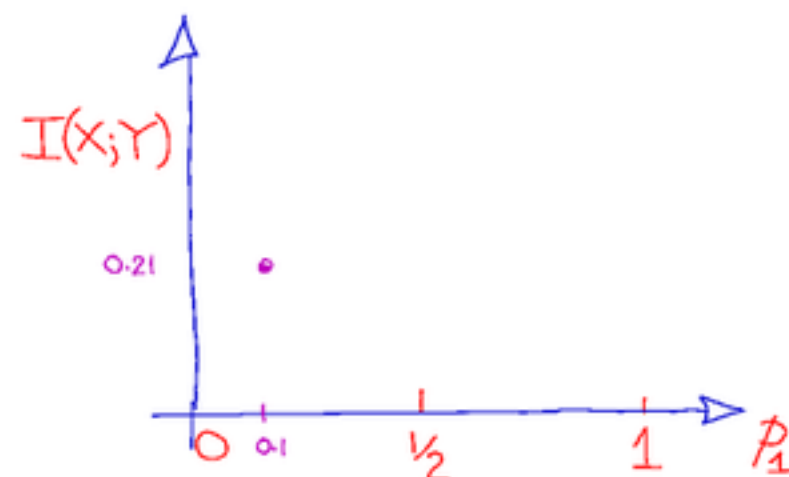
$$Q = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

Consider $f = 0.1$.

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

Find

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= H_2(0.18) - H_2(0.1) \\ &= 0.68 - 0.47 \\ &= 0.21 \text{ bits.} \end{aligned}$$



Complete the curve

Inference for channels

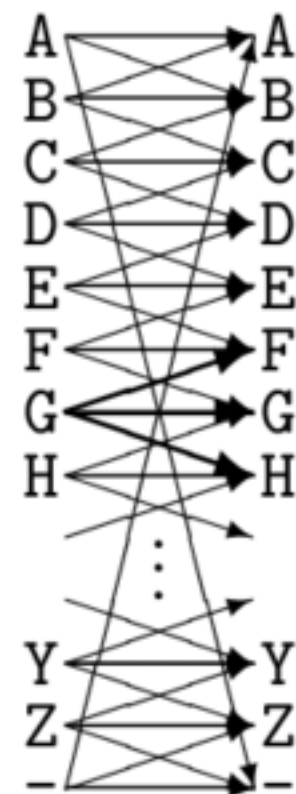
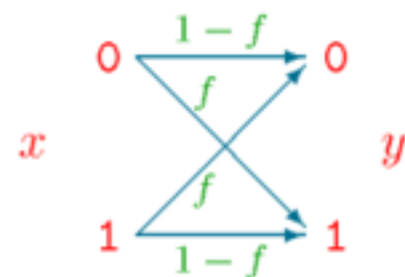
Channel Q defines conditional probabilities

$$P(y | x)$$

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

$$P(x, y) = P(x) P(y | x)$$

with which we can do inference.



Information measures for noisy channels

Channel Q defines conditional probabilities

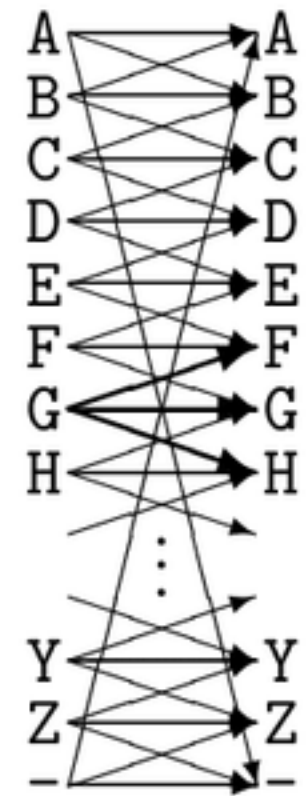
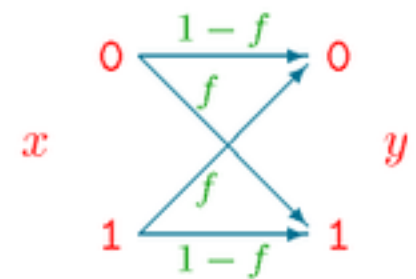
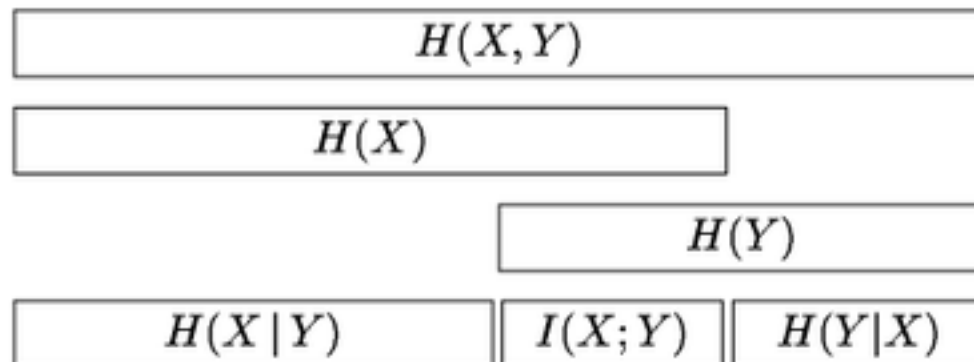
$$P(y | x)$$

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

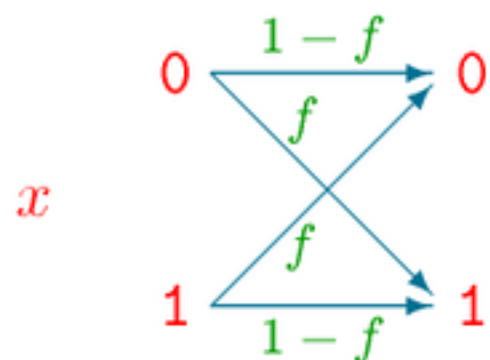
$$P(x, y) = P(x) P(y | x)$$

for which we can compute the mutual information

$$\begin{aligned} I(X; Y) &= H(X) - H(X | Y) \\ &= H(Y) - H(Y | X) \end{aligned}$$



Mutual information for the BSC



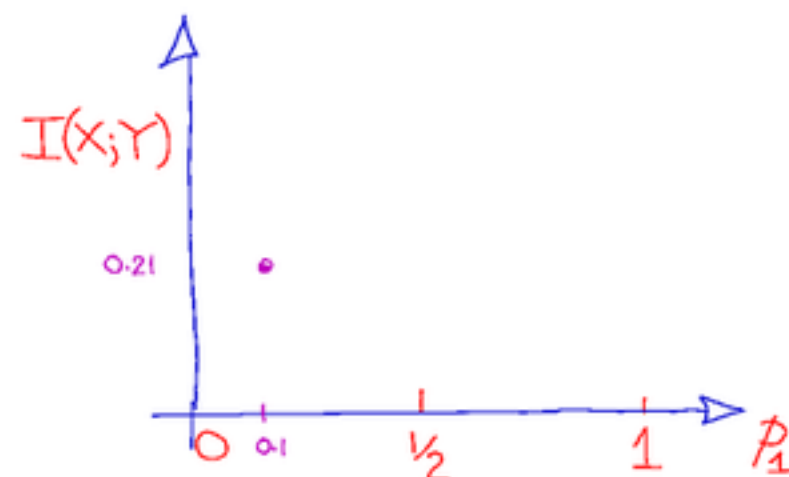
$$Q = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

Consider $f = 0.1$.

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

Find

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= H_2(0.18) - H_2(0.1) \\ &= 0.68 - 0.47 \\ &= 0.21 \text{ bits.} \end{aligned}$$



Complete the curve

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