Noisy Channels

Inference

& Information measures for noisy channels

Q: If we encode symbols from the ensemble

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

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

using the symbol code

$$\mathcal{C} = \{\mathtt{0}, \mathtt{10}, \mathtt{110}, \mathtt{111}\},$$

what is the probability 'p_1'
that a bit plucked at random from the encoded stream is a $\mathtt{1}$?

A $p_1 < \frac{1}{2}$

B $p_1 = \frac{1}{2}$

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

h_i	P_i			f_i	l_i
1	1/2	a	0	0	1
2	1/4	b	10	1/2	2
3	1/8	c	110	2/3	3
3	1/8	d	111	1	3

A	4
B	1
C	1
Z	

$$= \sum_i p_i f_i$$

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

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

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

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

$$7/4$$

$$\frac{\tfrac{1}{2}\times 0 + \tfrac{1}{4}\times 1 + \overbrace{\tfrac{1}{8}\times 2}^{1/4} + \overbrace{\tfrac{1}{8}\times 3}^{3/8}}{7/4} = \frac{7/8}{7/4} = 1/2$$

Last time - symbol codes

The ideal codelengths l_i^* are the information contents

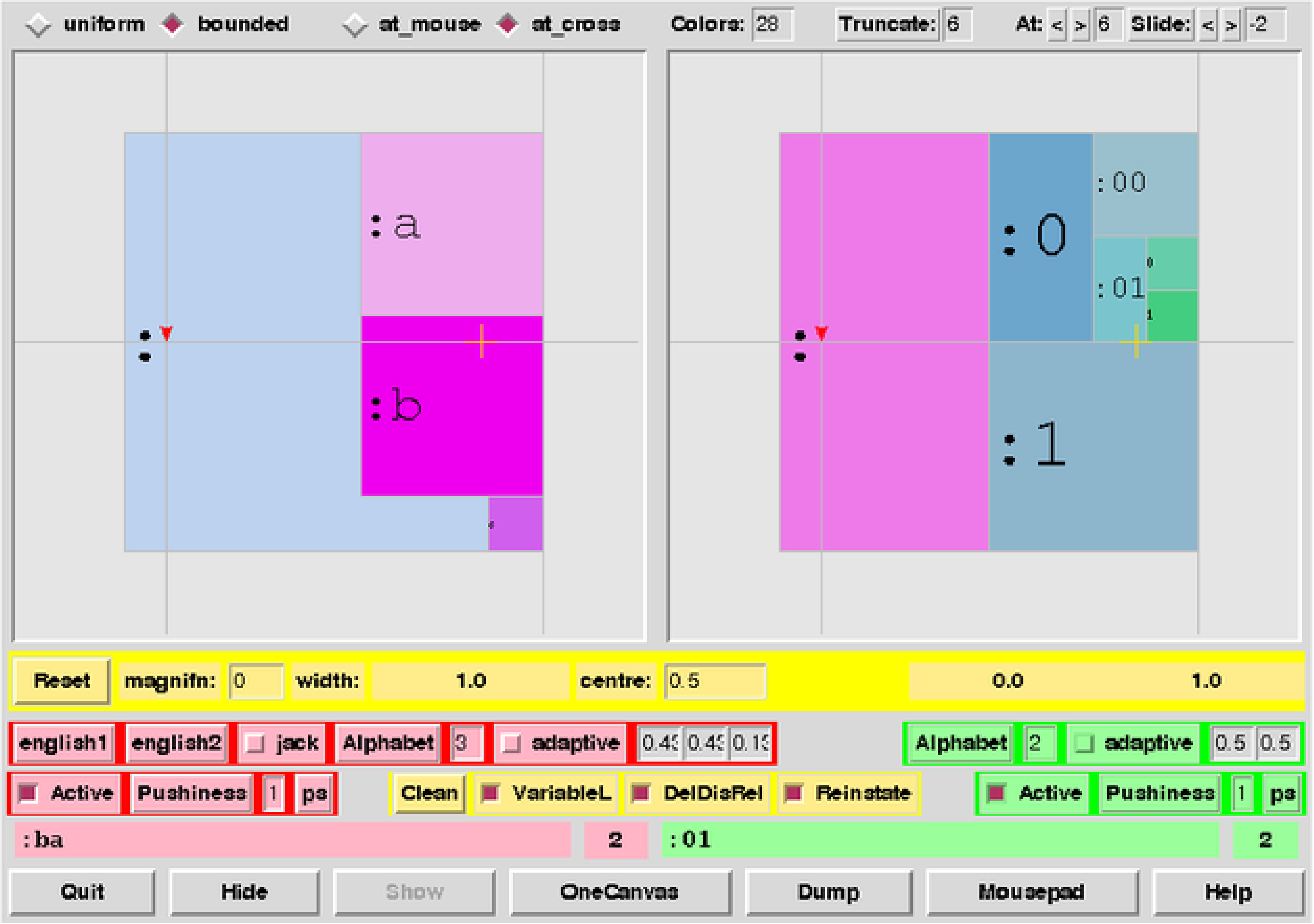
$$l_i^* = \log \frac{1}{p_i}$$

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



Arithmetic coding

achieves

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

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

Data compression

- An example predictor

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

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

Data compression

- An example predictor

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

```
..`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 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

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

```
`I'm afraid I don't know one,' said
Harriet, rather alarmed at the_pro
                                 e_pro
                                  _pro
                                   pro
                                    ro
                                     o
```

Other uses for arithmetic coding

- Efficient writing

Dasher
Information-Efficient Text Entry

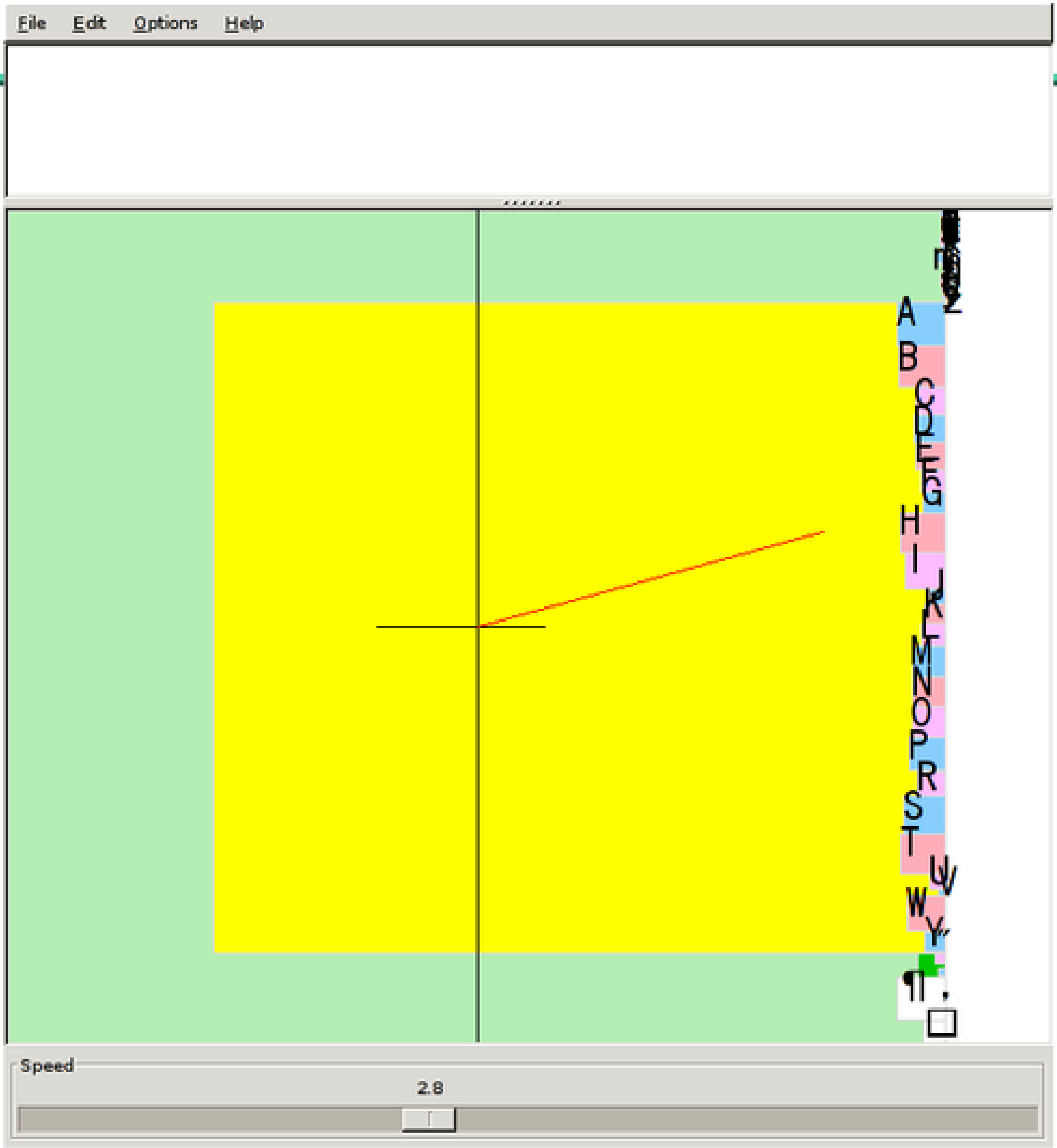
Compression:

Text $\longrightarrow$ Bit string
(preferably short)

Writing:

Text $\longleftarrow$ Gesture
(preferably brief)

Dasher



www.inference.phy.cam.ac.uk/dasher/

Other uses for arithmetic coding

Dasher
Information-Efficient Text Entry

- Efficient writing

Compression:

$$\text{Text} \longrightarrow \text{Bit string}$$

(preferably short)

Writing:

$$\text{Text} \longleftarrow \text{Gesture}$$

(preferably brief)

- Efficient generation of random samples

$P_a = 0.01$

$P_b = 0.99$

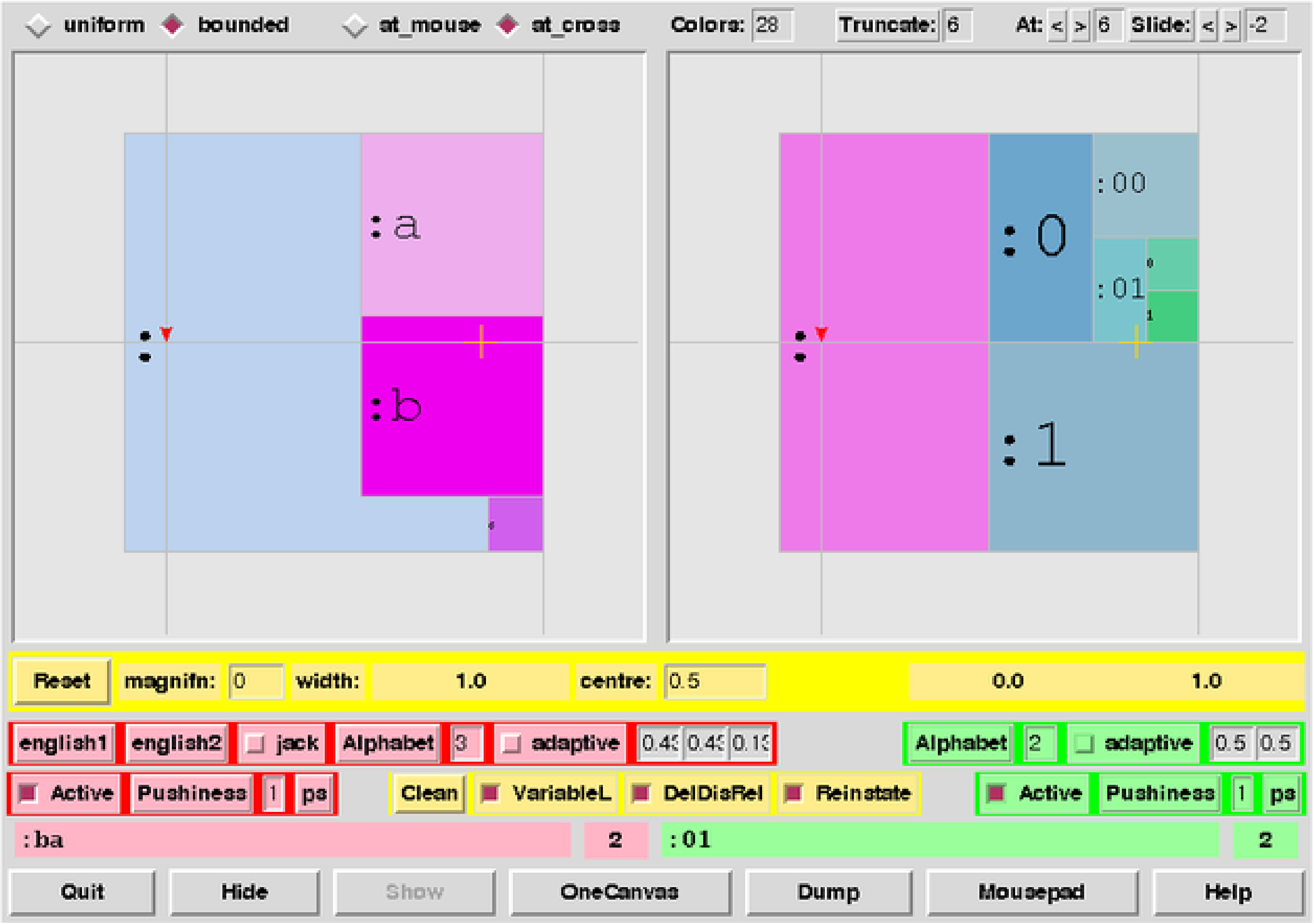
N

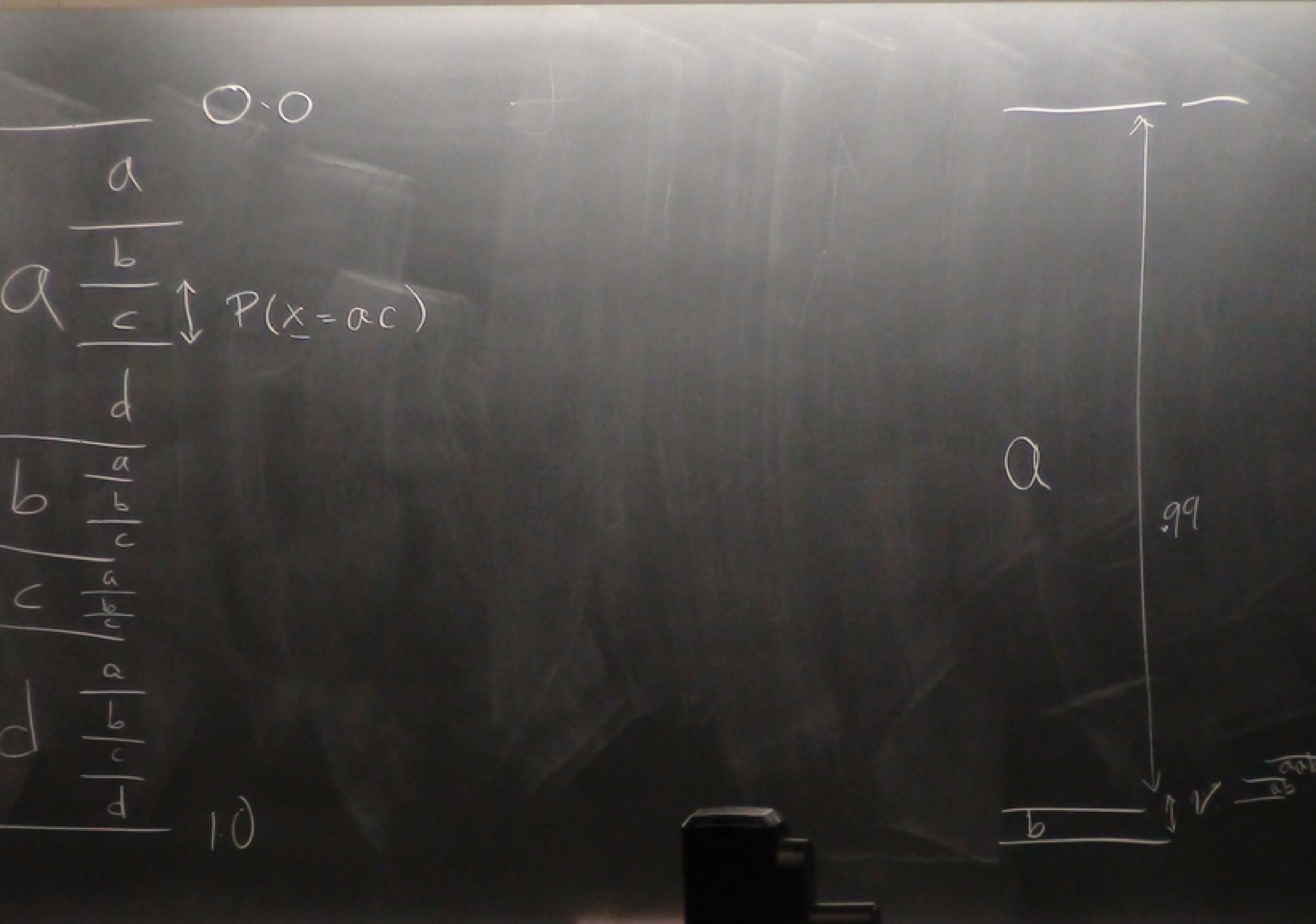
$u \sim (0,1)$

32 bits per real number

if $u < P_a \rightarrow a$

Arithmetic coding





0.0
a
b
c
d
P(x = ac)
a
b
c
d
a
b
c
d
a
b
c
d
a
b
c
d
1.0
a
.99
b

ac method

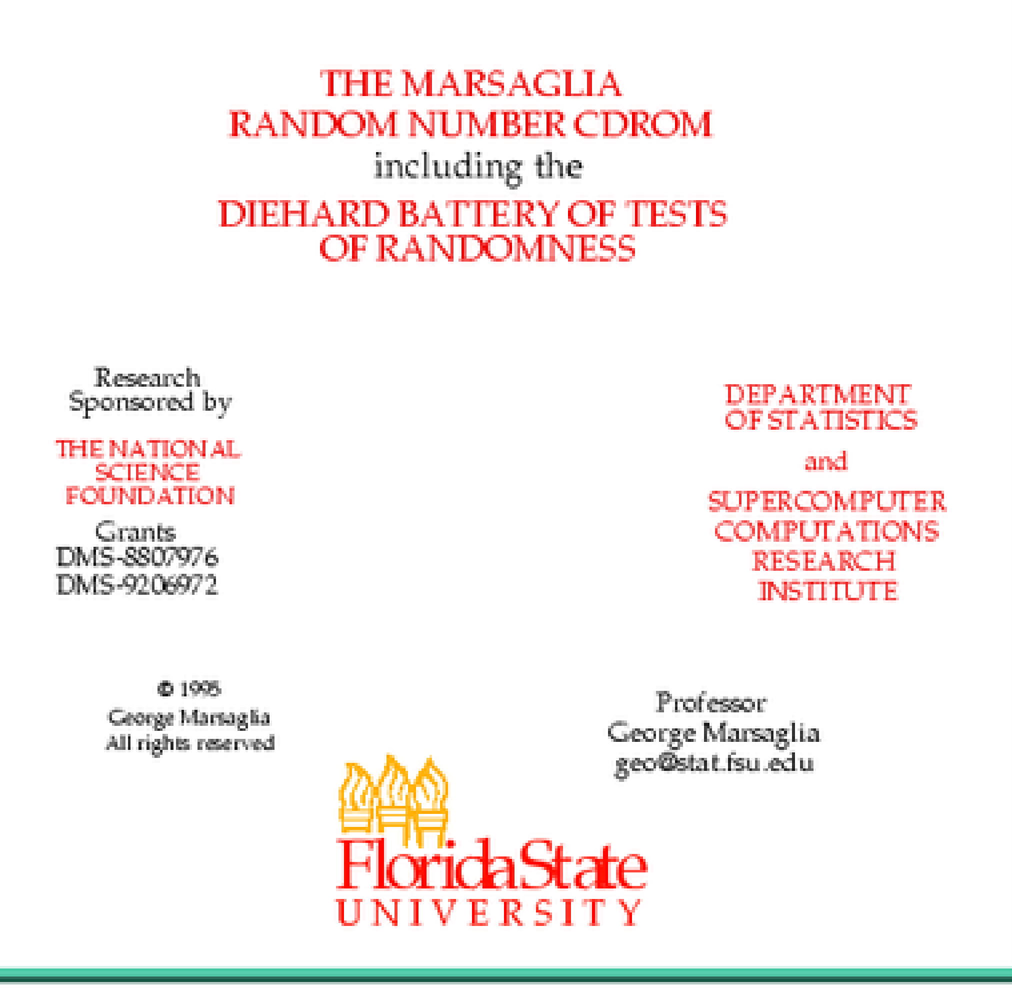
needs $H_2(0.01)$

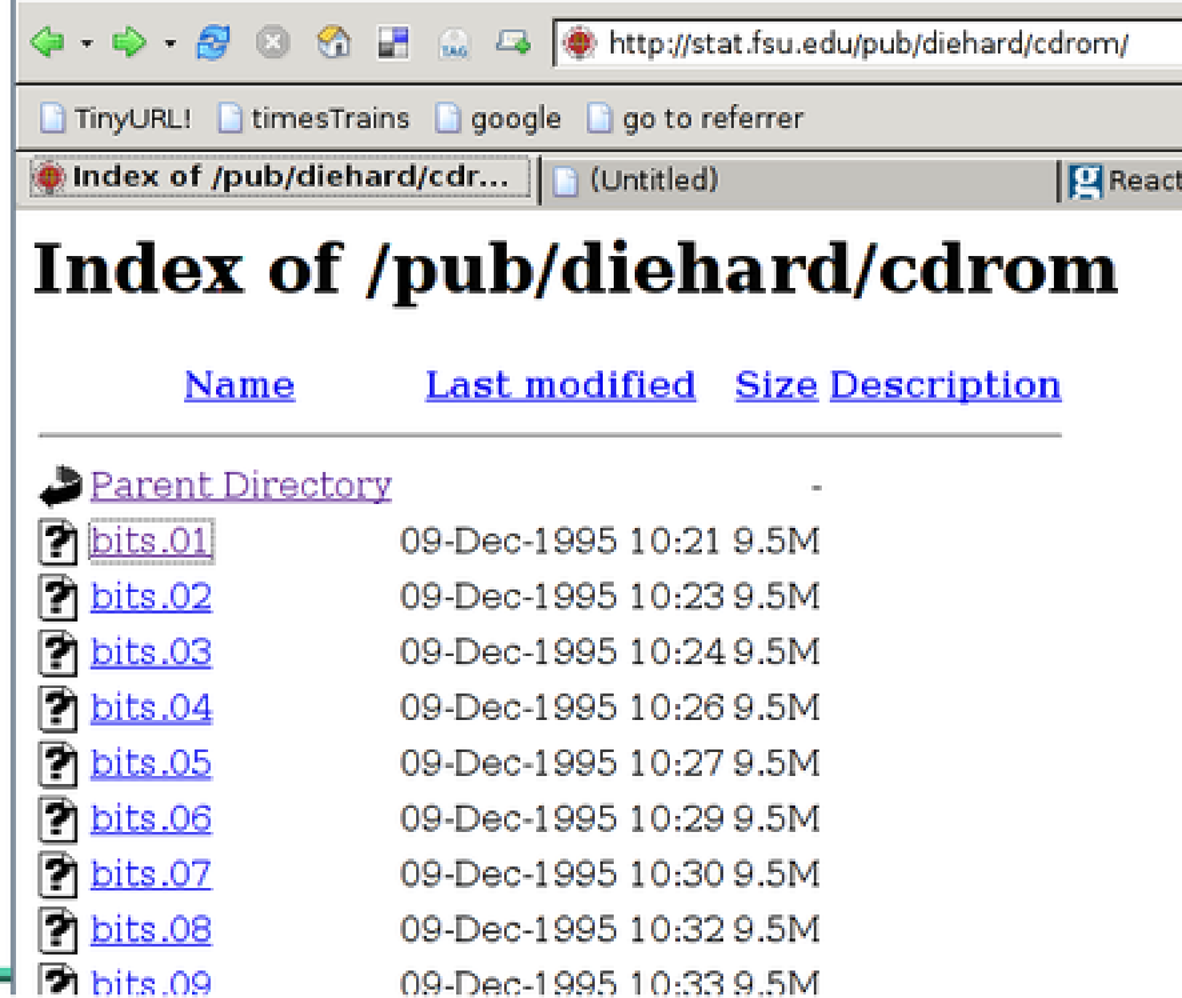
bits

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 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

$P_a = 0.01$

$P_b = 0.99$

N

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

$u \sim (0, 1)$

32 bits per real number

if $u < P_a \rightarrow a$

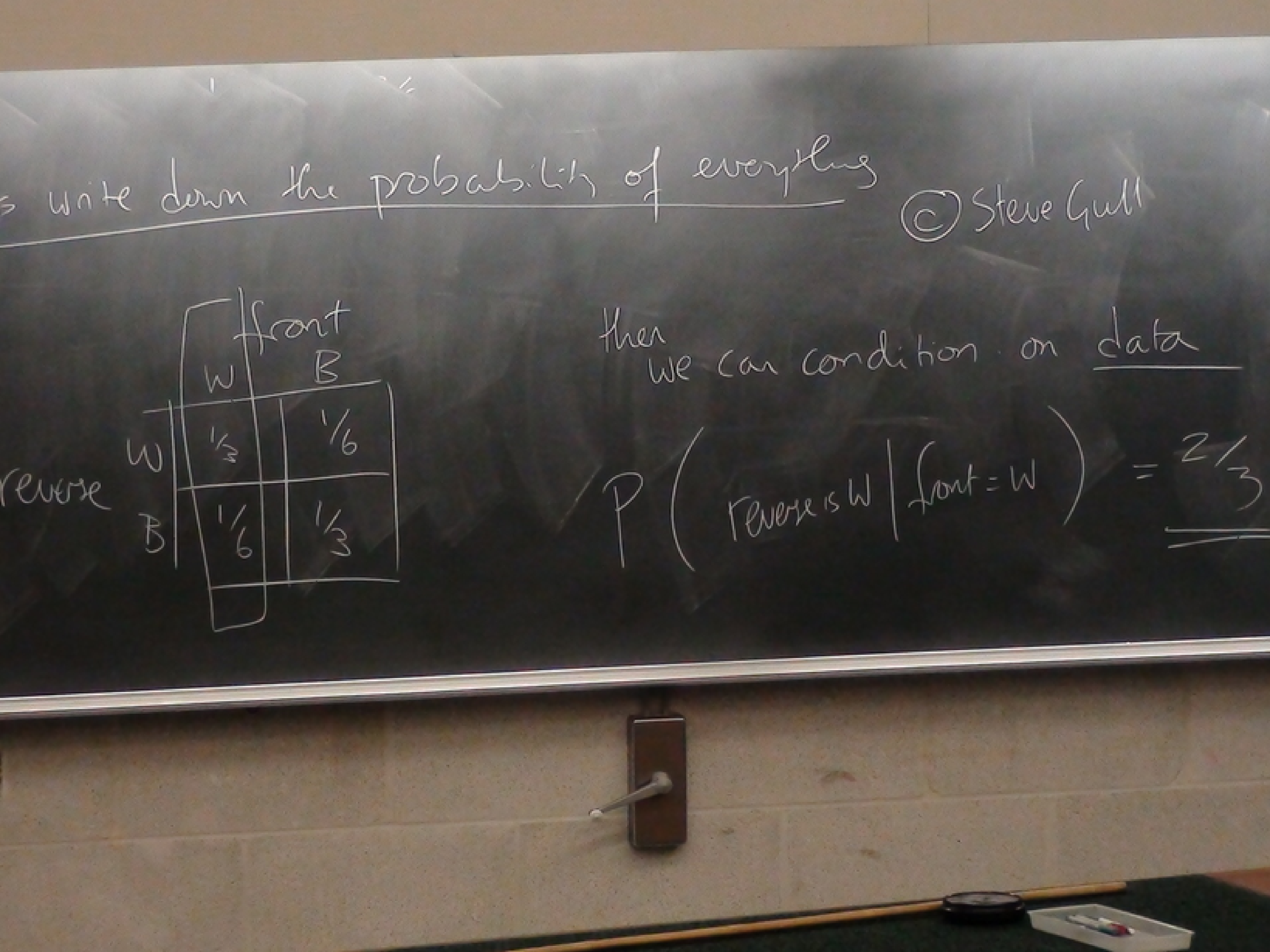
3 CARDS

W | W   B | B   B | W

P(other face is white | you see a white face uppermost) = ?

A	$p < 1/2$	0
B	$p > 1/2$	7
C	$p = 1/2$	6
Z	Don't know	1

Always write down the probability of everything



s write down the probability of everything
© Steve Gull
front
W
B
reverse
W
B
1/3
1/6
1/6
1/3
then we can condition on data
P (reverse is W | front = W) = 2/3

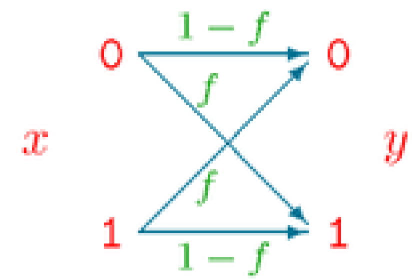
Three cards

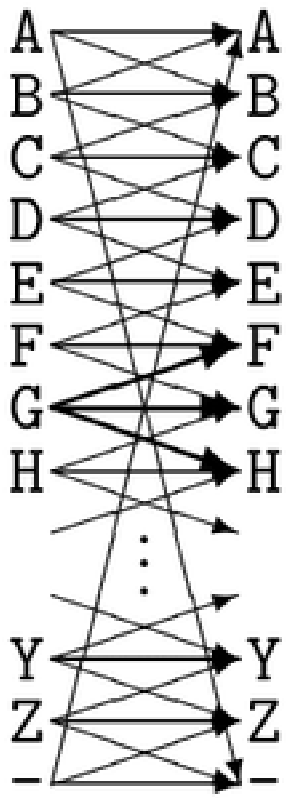
Q

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





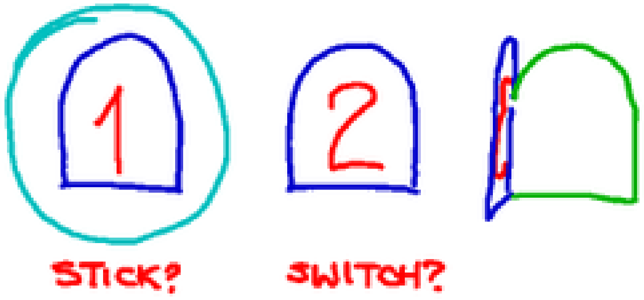
The rules

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



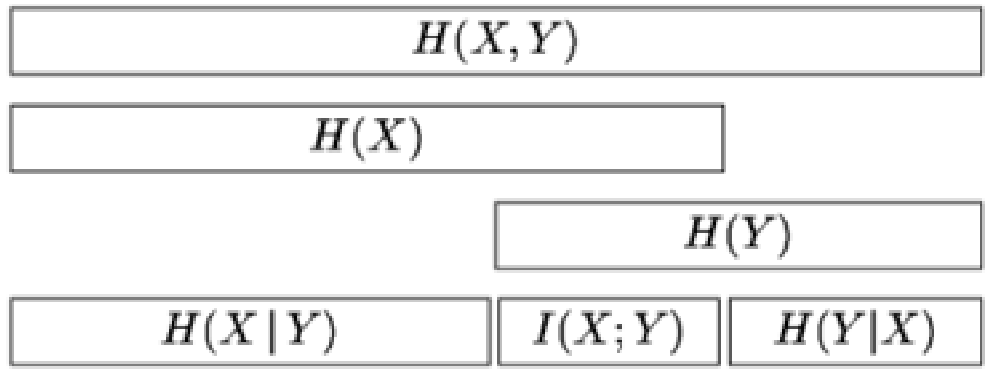
A should stick

B should switch

C makes no difference

Z don't know

Information measures



Example Joint ensemble

$$P(x,y)$$

Y \ X	1	2	3	4
1	$1/8$	$1/16$	$1/32$	$1/32$
2	$1/16$	$1/8$	$1/32$	$1/32$
3	$1/16$	$1/16$	$1/16$	$1/16$
4	$1/4$	0	0	0

nsemble

Marginal

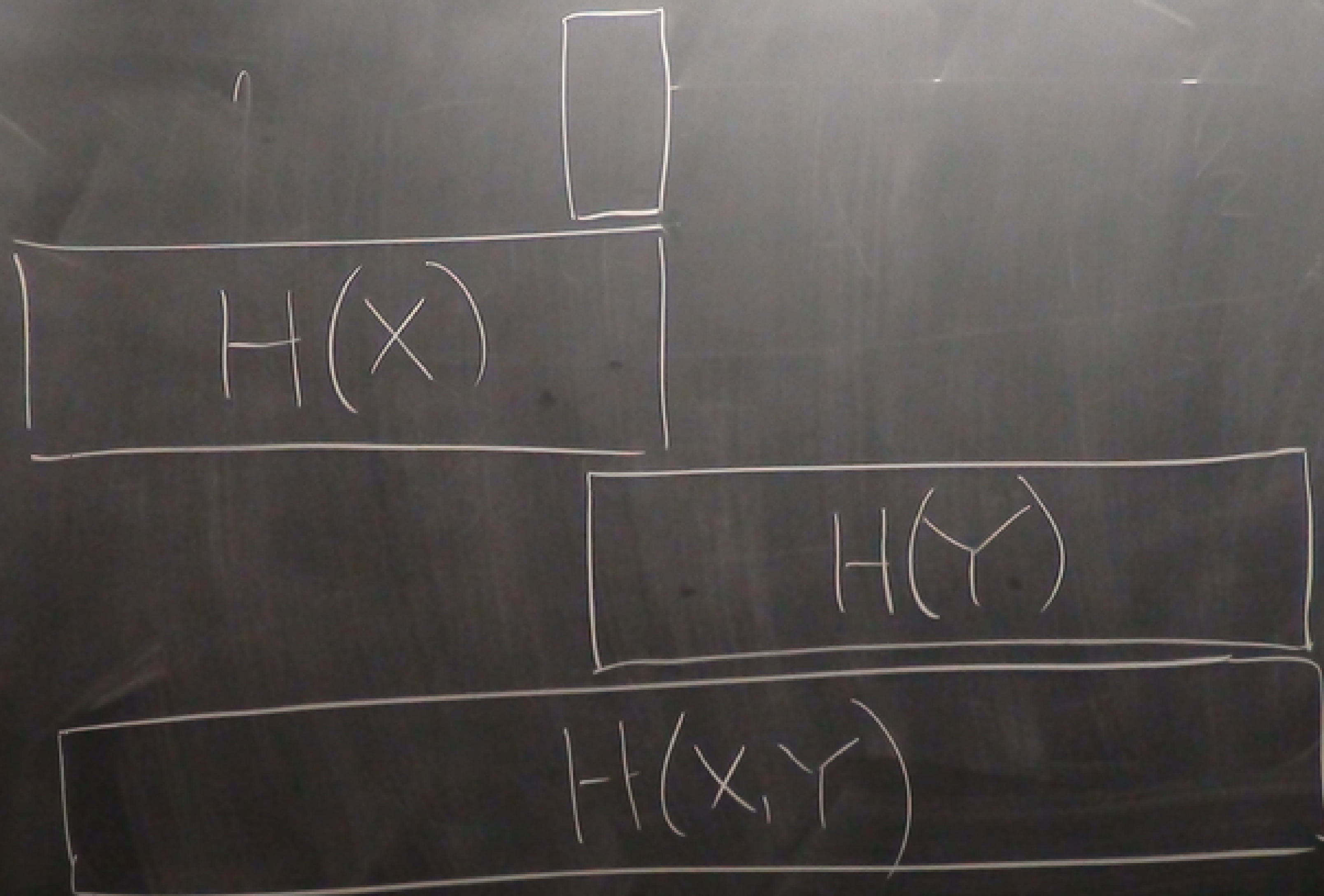
y \ x	1	2	3	4	$P(y)$
1	1/8	1/16	1/32	1/32	1/4
2	1/16	1/8	1/32	1/32	1/4
3	1/16	1/16	1/16	1/16	1/4
4	1/4	0	0	0	1/4
	[illegible]	1/4	1/8	1/8	

"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)
H(Y)
H(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 prob^y

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

$P(x|y)$

y \ x	1	2	3	4
1	1/2	1/4	1/8	1/8
2	1/4	1/2	1/8	1/8
3	1/4	1/4	1/4	1/4
4	1	0	0	0

$$H(X \mid y=1) = 7/4$$

$$H(X \mid y=2) = 7/4$$

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

$$= 0 \text{ bits}$$

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

CONDITIONAL ENTROPY

$$H(X|y=1) = 7/4$$

$$H(X|y=2) = 7/4$$

$$= 2 \text{ bits}$$

$$\vdots$$

$$= 0 \text{ bits}$$

$$H(X|Y) = \sum_y P(y)\, H(X|y)$$

$$= 11/8 \text{ bits}$$

Conditional prob^y

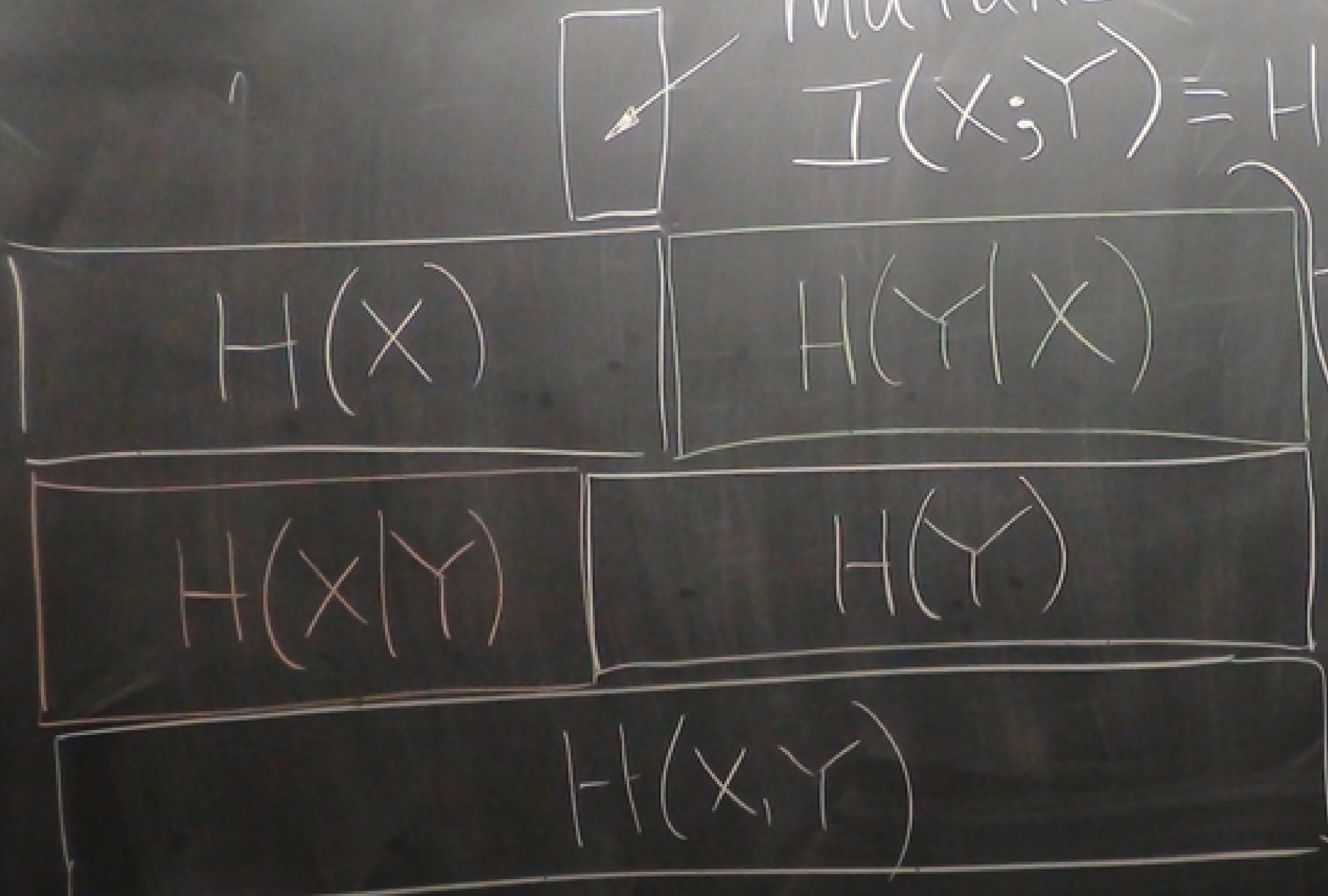
$$P(x|y) = \frac{P(x,y)}{P(y)}$$

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

MUTUAL INFORMATION

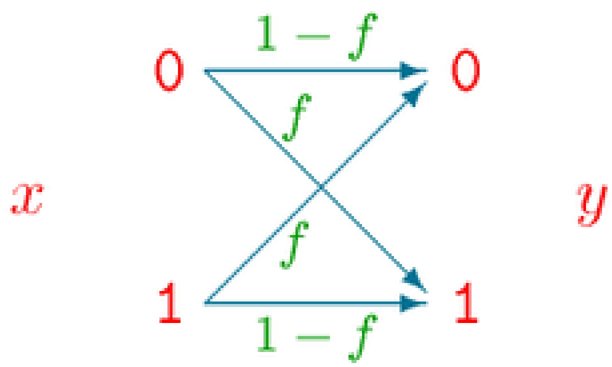
$I(X;Y) = H(X) - H(X|Y)$ (Condit

$= H(Y) - H(Y|X)$ P(x



$H(X|Y$

Mutual information for the BSC



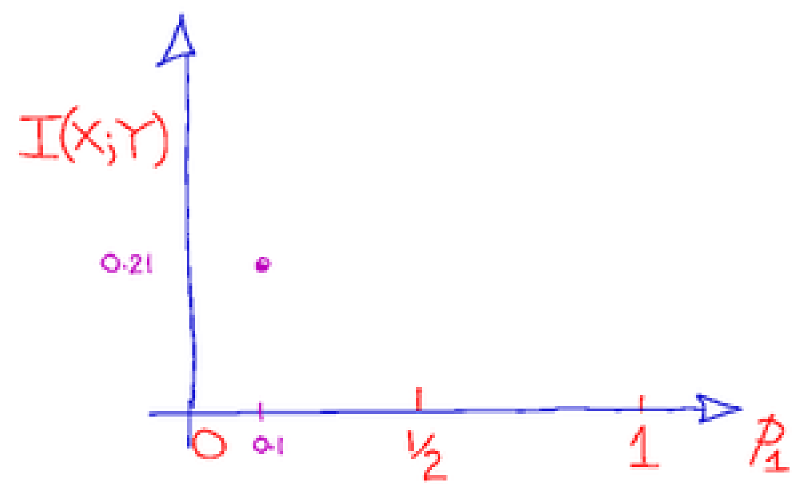
$$\mathbf{Q} = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

Consider $f = 0.1$.

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

Find

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



Complete the curve

Inference for channels

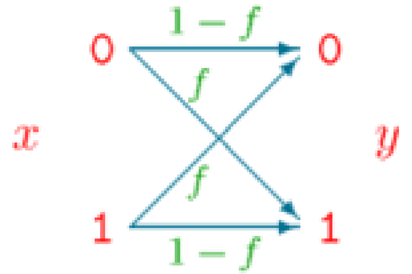
Channel **Q** defines conditional probabilities

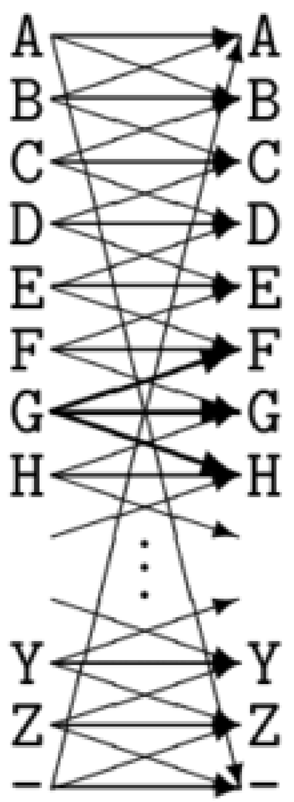
$$P(y \mid x)$$

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.

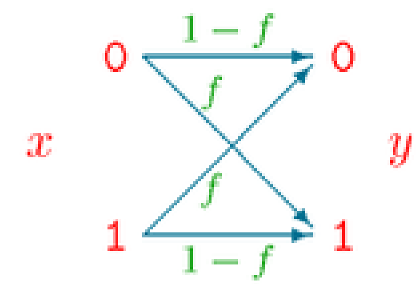




Information measures for noisy channels

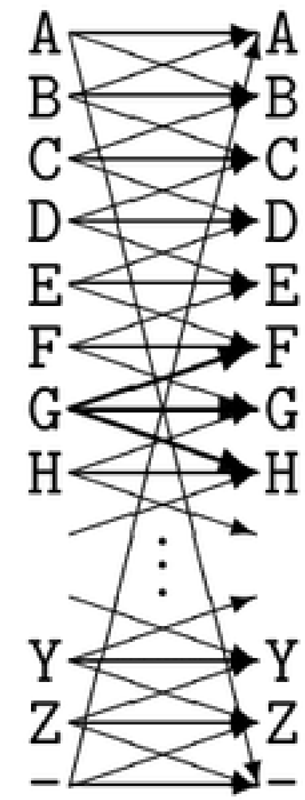
Channel **Q** defines conditional probabilities

$$P(y \mid x)$$



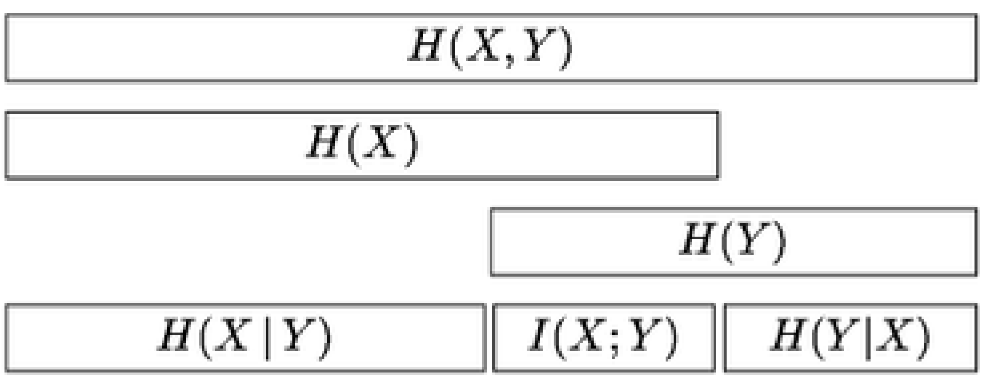
If we choose an input distribution $P(x)$,
we have a joint distribution

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

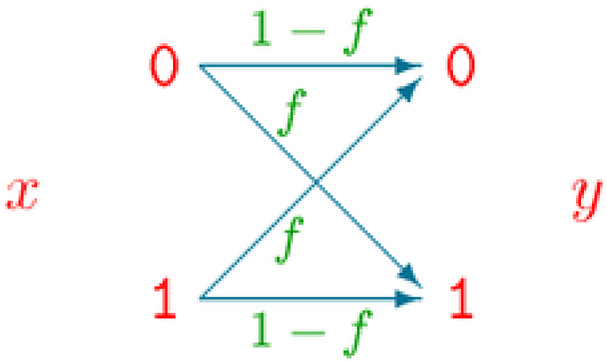


for which we can compute the mutual information

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



Mutual information for the BSC



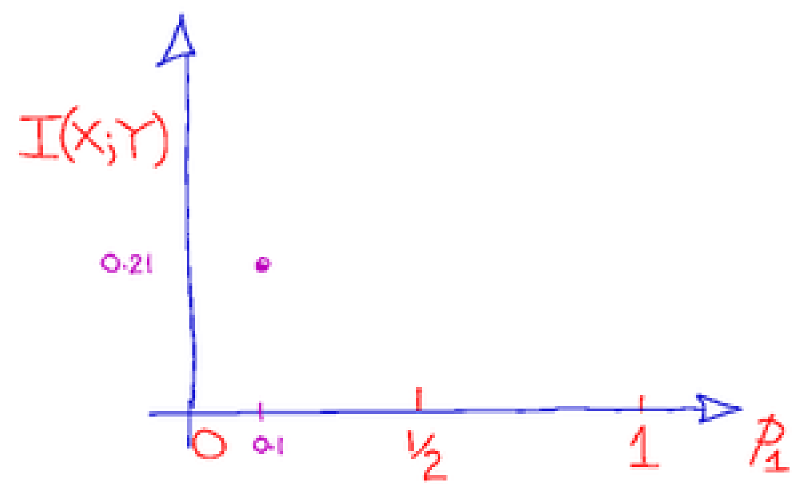
$$\mathbf{Q} = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

Consider $f = 0.1$.

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

Find

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



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