September 20, 2016
Tuesday

Information Measures:	What they measure
Entropy	Randomness, Information
Mutual Information	Correlation, Amount of information on R.V. contains about other.
Relative Entropy	"Distance" between distributions

Let X be a discrete R.V.

$\mathcal{X}$ is the alphabet

$P_X(x)$ is the prob. mass funct. (i.e. $P(X=x)=P_X(x)$)

$P_X(x) \geq 0$

$\sum_{x\in\mathcal{X}} P_X(x) = 1$

Entropy: $H(X) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{1}{P_X(x)} = -\sum_{x\in\mathcal{X}} P_X(x) \log P_X(x)$

Okay to write $H(P_X)$ instead

Expected Value: $\mathbb{E}[g(X)] = \sum_{x\in\mathcal{X}} P_X(x) g(x)$

$$H(X) = \mathbb{E}\left[\log \frac{1}{P_X(X)}\right]$$

Convention: $0 \log 0 = 0$ (or $0 \log \infty = 0$)

because $\lim_{x\to 0} x \log x = 0$

Properties:

1) $H(X) \geq 0$

proof: $\log \frac{1}{P_X(x)} \geq 0$ w.p. 1 $\quad P_X(x) \in [0,1] \; \forall x$

2) Base of log gives units

$$H_b(X) = \mathbb{E}\left[\log_b \frac{1}{P_X(X)}\right]$$

$$H_b(X) = (\log_b a)\, H_a(X)$$

↑ unit conversions

base	Units
2	bits
e	nats
10	decimal digit

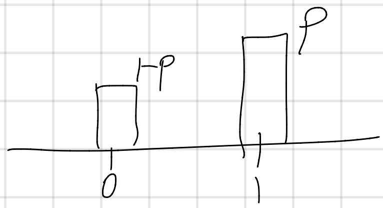
← binary digit

3) $H(X) \leq \log |\mathcal{X}|$

Example 1

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

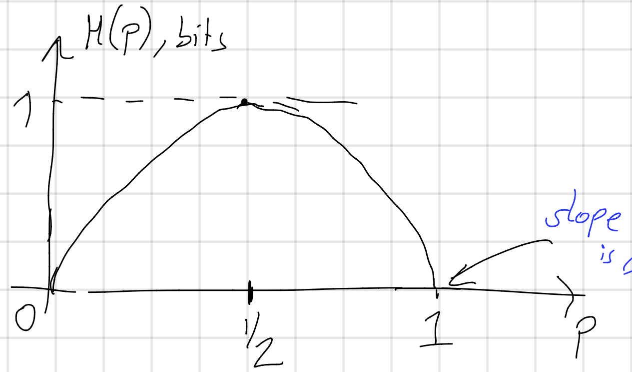
$X \sim Ber(p)$



$$H(X) = (1-p)\log\frac{1}{1-p} + p\log\frac{1}{p}$$

$H(p) \triangleq$

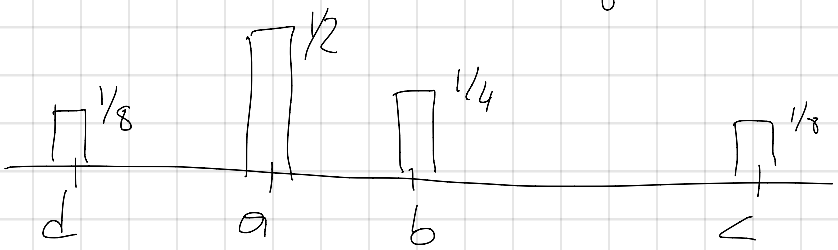
"binary entropy function"



$H(X) = 0$ iff X is deterministic

Example 2:

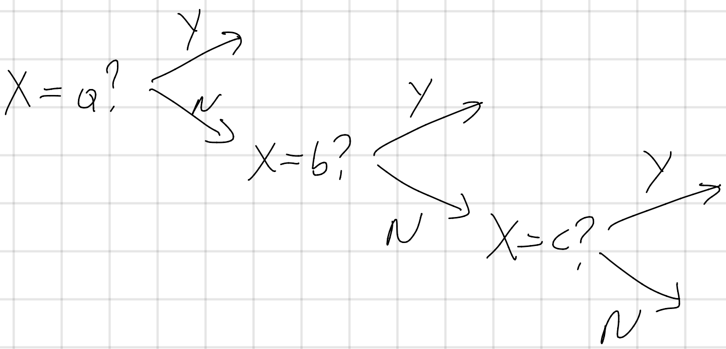
$$X = \begin{cases} a & \text{w.p. } 1/2 \\ b & 1/4 \\ c & 1/8 \\ d & 1/8 \end{cases}$$



$$H(X) = \frac{1}{2}\log\frac{1}{1/2} + \frac{1}{4}\log\frac{1}{1/4} + \frac{1}{8}\log\frac{1}{1/8} + \frac{1}{8}\log\frac{1}{1/8}$$

$$= \frac{1}{2}\cdot 1\text{ bit} + \frac{1}{4}\cdot 2\text{ bits} + \frac{1}{4}\cdot 3\text{ bits}$$

$$= 1.75$$



	# of questions
a	1
b	2
c	3
d	3

$$\mathbb{E}[\#\text{ of questions}] = 1.75$$

Code:

a	1
b	01
c	001
d	000

$\rightsquigarrow$ acabd → 1001101000

↳ prefix code

$$\mathbb{E}[\# \text{ of Y/N questions}] \in [H(X), H(X)+1] \text{ bits}$$

(Note that if $g(\cdot)$ is invertible function and $Y = g(X)$
$H(Y) = H(X)$)

Joint Entropy:

$$H(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \log \frac{1}{P_{XY}(x,y)} = \mathbb{E}\left[\log \frac{1}{P_{XY}(X,Y)}\right]$$

Conditional Entropy

$$H(X|Y) = \mathbb{E}\left[\log \frac{1}{P_{Y|X}(Y|X)}\right] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \log \frac{1}{P_{Y|X}(y|x)}$$

Chain Rule: $H(X,Y) = H(X) + H(Y|X)$

Proof: $\curvearrowright$ $= \mathbb{E}\left[\log \frac{1}{P_X(X)}\right] + \mathbb{E}\left[\log \frac{1}{P_{Y|X}(Y|X)}\right]$

$$= \mathbb{E}\left[\log \frac{1}{P_X(X) P_{Y|X}(Y|X)}\right]$$

Comment on Conditional Entropy:

$$H(P_{Y|X=x}) = \sum_y P_{Y|X=x}(y) \log \frac{1}{P_{Y|X=x}(y)}$$

$$H(Y|X) = \sum_{x \in \mathcal{X}} P_X(x) H(P_{Y|X=x})$$

Example:

$Y \backslash X$	1	2	3	4
	$1/8$	$1/16$	$1/32$	$1/32$
	$1/16$	$1/8$	$1/32$	$1/32$
	$1/16$	$1/16$	$1/16$	$1/16$
	$1/4$	0	0	0
P_X	$\frac{1}{2}$	$1/4$	$1/8$	$1/8$

$$H(X,Y) = \frac{27}{8} \text{ bits}$$

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

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

$$H(P_{X|Y=1}) = \frac{7}{4} \text{ bits} \qquad H(P_{X|Y=2}) = \frac{7}{4} \text{ bits}$$

$$H(P_{X|Y=3}) = 2 \text{ bits} \qquad H(P_{X|Y=4}) = 0 \text{ bits}$$

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

$$H(X|Y) = H(X,Y) - H(Y) = \frac{11}{8} \text{ bits}$$

Relative entropy (expected log-likelyhood)

$$D(P\|Q) = \mathbb{E}\left[\log \frac{P(X)}{Q(X)}\right] \quad \text{where } X \sim P$$

$D(P\|Q)$ is the ineff. of encoding under the assumption Q when true distribution is P.

distributions

$$D(P\|Q) = \mathbb{E}_P\left[\log \frac{P(X)}{Q(X)}\right] \quad \text{when } X \sim P$$

$$= \sum_{x \in \mathcal{X}} P_X(x) \log \frac{P(x)}{Q(x)}$$

If $\exists x$ s.t. $Q(x) = 0$ when $P(x) > 0 \Rightarrow D(P\|Q) = \infty$

(In general, $D(P\|Q) = \mathbb{E}\left[\log \frac{dP}{dQ}(X)\right]$ where $X \sim P$ and $P \ll Q$

If $P \not\ll Q$ then $D(P\|Q) = \infty$)

Mutual Information

$$I(X;Y) = D(P_{XY} \| P_X P_Y) = \mathbb{E}\left[\log \frac{P_{XY}(X,Y)}{P_X(X) P_Y(Y)}\right]$$

when X and Y are <u>discrete</u> r.v.'s

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

why?

$$I(X;Y) = \mathbb{E}\left[\log \frac{P_{Y|X}(Y|X)}{P_Y(Y)}\right]$$

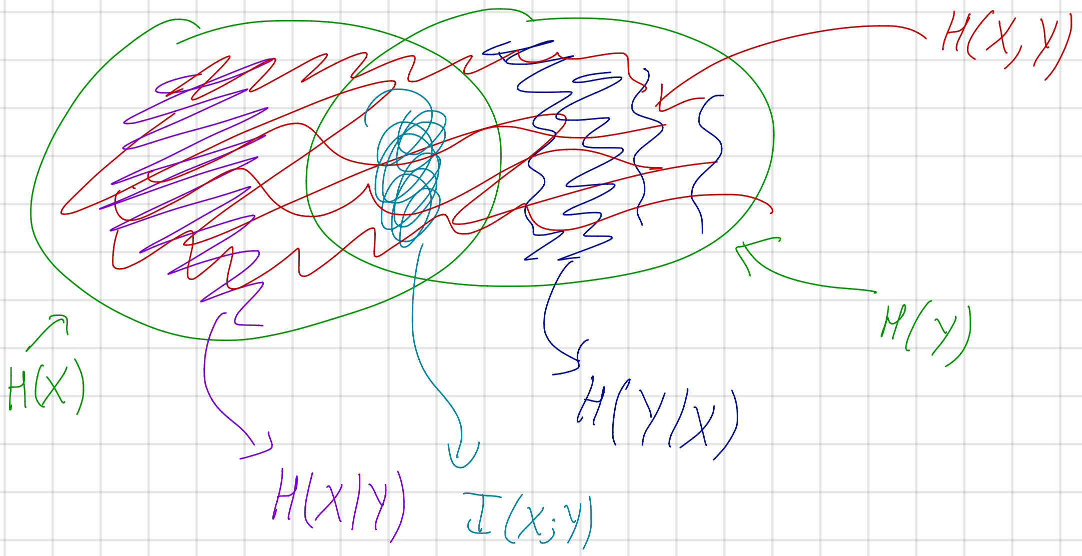
$$= \mathbb{E}\left[\log \frac{1}{P_Y(Y)}\right] - \mathbb{E}\left[\log \frac{1}{P_{Y|X}(Y|X)}\right]$$

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

$$= H(X) + H(Y) - H(X,Y)$$

$$= I(Y;X)$$

$$I(X;X) = H(X)$$



Example:

Y \ X	1	2	3	4
	1/8	1/16	1/32	1/32
	1/16	1/8	1/32	1/32
	1/16	1/16	1/16	1/16
	1/4	0	0	0
P_X	$\frac{1}{2}$	1/4	1/8	1/8

$$H(X,Y) = \frac{27}{8} \text{ bits}$$

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

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

$$\boxed{I(X;Y) = 0.375 \text{ bits}}$$

Conditional mutual information

$$I(X;Y|Z) = \mathbb{E}\left[\log \frac{P_{XY|Z}(X,Y|Z)}{P_{X|Z}(X|Z)\, P_{Y|Z}(Y|Z)}\right] = D\left(P_{XY|Z} \,\|\, P_{X|Z} P_{Y|Z} \,|\, P_Z\right)$$

Conditional Relative Entropy

$$D(P_{X|Y} \| Q_{X|Y} | P_Y) = \mathbb{E}_{P_Y P_{X|Y}}\left[\log \frac{P_{X|Y}(X|Y)}{Q_{X|Y}(X|Y)}\right]$$

Prop

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z)$$

Chain Rules:

$$H(X_1,\dots,X_n) = \sum_{i=1}^{n} H(X_i|X_1 \cdots X_{i-1})$$

$$I(X_1,\dots,X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_1,\dots,X_{i-1})$$

$$D(P_{XY} \| Q_{XY}) = D(P_X \| Q_X) + D(P_{Y|X} \| Q_{Y|X} | P_X)$$