RWTH AACHEN UNIVERSITY

La. 2.2.5. If X and Y are discrete i.i.d. r.v. with entropy $H(X)$. Then
$$P(X=Y) \geq 2^{-H(X)}.$$

Proof. Let $p(x)$ denote the pmf of X and Y.
W.l.o.g. choose log-base $=2$.
$f(t) = 2^t$ is a convex fct. Hence,
$$\begin{aligned} 2^{-H(X)} &= 2^{E(\log p(X))} \\ &\leq E\left(2^{\log p(X)}\right) \quad \text{(by Jensen inequ.)} \\ &= \sum_x p(x)\, 2^{\log p(x)} \\ &= \sum_x p^2(x) = P(X=Y). \quad \boxtimes \end{aligned}$$

Is this inequ. sharp?

a) Let X, Y i.i.d. $\sim U(1, \dots, m)$ (uniformly distributed)
$$\left.\begin{aligned} H(X) &= \log m, \quad 2^{-H(X)} = \frac{1}{m} \\ P(X=Y) &= \sum_{i=1}^{m} \frac{1}{m^2} = \frac{1}{m} \end{aligned}\right\} =$$

b) X, Y i.i.d. with only one mass point.
$$\left.\begin{aligned} &H(X) = 0, \quad 2^{-H(X)} = 1 \\ &P(X=Y) = 1 \end{aligned}\right\} =$$

Inequality is sharp at least for two cases.

Thinking the Future
Zukunft denken

2.3. Information Measures for Random Sequences

Consider sequences of r.v. $X_1, X_2, X_3, \dots$ denoted as $X = \{X_n\}_{n\in\mathbb{N}}$.

Naive approach: define the entropy of X

$$H(X) = \lim_{n\to\infty} H(X_1, X_2, \dots, X_n).$$

In most cases this limit will be infinite.

Instead consider the <u>entropy rate</u>.

<u>Def. 2.3.1.</u> Let $X = \{X_n\}_{n\in\mathbb{N}}$ be a sequence of discrete r.v.

$$H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \dots, X_n)$$

is called the <u>entropy rate</u> of X, provided the limit exists. ⌋

$H_\infty(X)$ is the average uncertainty per symbol.

<u>Def. 2.3.2.</u> Let $X = \{X_n\}_{n\in\mathbb{N}}$, $Y = \{Y_n\}_{n\in\mathbb{N}}$ sequ. of discr. r.v.

$$I_\infty(X, Y) = \lim_{n\to\infty} \frac{1}{n} I(X_1, \dots, X_n; Y_1, \dots, Y_n)$$

is called <u>mutual information rate</u> of X and Y.

<u>Example 2.3.2</u>

a) Let $X = \{X_n\}_{n\in\mathbb{N}}$ be i.i.d. r.v. with $H(X_i) < \infty$.

Then

$$H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \dots, X_n)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) = H(X_1)$$

b) Let $\{Z_n\}_{n\in\mathbb{N}} = \{(X_n, Y_n)\}_{n\in\mathbb{N}}$ be i.i.d sequence with $I(X_k; Y_k) < \infty$. Then

$$I_\infty(X;Y) = \lim_{n\to\infty} \frac{1}{n} I(X_1,\dots,X_n; Y_1,\dots,Y_n)$$

$$\overset{(Ex.)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(X_k; Y_k)$$

$$= I(X_1; Y_1) \qquad \blacksquare$$

Going further than i.i.d. sequences.

<u>Def. 2.3.3.</u> A sequence of r.v.s $X = \{X_n\}_{n\in\mathbb{N}}$ is called (strongly) <u>stationary</u> if

$$P^{(X_{i_1},\dots,X_{i_k})} = P^{(X_{i_1+t},\dots,X_{i_k+t})}$$

for all $1 \le i_1 < \dots < i_k$, $t \in \mathbb{N}$.

- The joint distribution of any finite selection of r.v.s from $\{X_n\}_{n\in\mathbb{N}}$ is invariant w.r.t. time shift.
- An equivalent condition for discrete r.v. with support $\mathcal{X}$ is as follows

$$P(X_1 = s_1, \dots, X_n = s_n) = P(X_{1+t} = s_1, \dots, X_{n+t} = s_n)$$

for all $s_1,\dots,s_n \in \mathcal{X}$, $n\in\mathbb{N}$, $t\in\mathbb{N}$.

- For stationary sequences all marginal distributions P^{X_ℓ} are the same

Theorem 2.3.4. Let $X = \{X_n\}_{n \in \mathbb{N}}$ be a stationary sequ. Then

a) $H(X_n | X_1, \ldots, X_{n-1})$ is monotonically decreasing.

b) $H(X_n | X_1, \ldots, X_{n-1}) \leq \frac{1}{n} H(X_1, \ldots, X_n)$

c) $\frac{1}{n} H(X_1, \ldots, X_n)$ is monotonically decreasing.

d) $\lim\limits_{n \to \infty} H(X_n | X_1, \ldots, X_{n-1}) = \lim\limits_{n \to \infty} \frac{1}{n} H(X_1, \ldots, X_n) = H_\infty(X)$

Because of a) and c) both limits exist. ⌟

Proof.

a) $H(X_n | X_1, \ldots, X_{n-1}) \underset{\text{Th. 2.1.8 d)}}{\leq} H(X_n | X_2, \ldots, X_{n-1}) \underset{\text{stationary}}{\leq} H(X_{n-1} | X_1, \ldots, X_{n-2})$

Hence $H(X_n | X_1, \ldots, X_{n-1})$ is mon. decreasing.

The first limit in d) exists, since 0 is a lower bound.

b) By Th. 2.1.5 (chain rule)

$$\frac{1}{n} H(X_1, \ldots, X_n) = \frac{1}{n}\big(H(X_1) + H(X_2 | X_1) + \cdots + H(X_n | X_1, \ldots, X_{n-1})\big)$$

$$\overset{a)}{\cancel{\leq H(X}}$$

$$\geq H(X_n | X_1, \ldots, X_{n-1})$$

c) $H(X_1, \ldots, X_n) = H(X_1, \ldots, X_{n-1}) + H(X_n | X_1, \ldots, X_{n-1})$

$$\overset{b)}{\leq} H(X_1, \ldots, X_{n-1}) + \frac{1}{n} H(X_1, \ldots, X_n)$$

Hence,

$$\frac{n-1}{n} H(X_1, \ldots, X_n) \leq H(X_1, \ldots, X_{n-1})$$

$$\Leftrightarrow \quad \frac{1}{n} H(X_1, \ldots, X_n) \leq \frac{1}{n-1} H(X_1, \ldots, X_{n-1})$$

which proves monotonicity.

d) $$\frac{1}{n+k} H(X_1,\dots,X_{n+k})$$

$$= \frac{1}{n+k}\Big[H(X_{n+k} \mid X_1,\dots,X_{n+k-1}) + \dots + H(X_{n+1} \mid X_1,\dots,X_n)$$

$$+ H(X_n \mid X_1,\dots,X_{n-1}) + H(X_1,\dots,X_{n-1}) \Big]$$

$$\le \underbrace{\frac{1}{n+k} H(X_1,\dots,X_{n-1})}_{\to 0 \ (k\to\infty)} + \underbrace{\frac{k+1}{n+k} H(X_n \mid X_1,\dots,X_{n-1})}_{\to H(X_n \mid X_1,\dots,X_{n-1}) \ (k\to\infty)}$$

Now fix n, set $\ell = n+k$

Using b)

$$\lim_{n\to\infty} H(X_n \mid X_1,\dots,X_{n-1}) \le \lim_{n\to\infty} \frac{1}{n} H(X_1,\dots,X_n)$$

$$\le \lim_{n\to\infty} H(X_n \mid X_1,\dots,X_{n-1})$$

so that the limits ~~e~~ exist and are equal. $\square$

RWTH AACHEN UNIVERSITY

Example 2.3.5. (English text)

Frequencies of single characters

A	B	..	E	..	Z
0.082	0.015		0.127		0.001

Frequencies of digrams (in %)

AN	TH	..	TI
1.81	3.21		1.28

Estimated entropy rates (log-base = 2)

n	1	2	3	..
$\frac{1}{n} H(X_1, \dots, X_n)$	4.14	3.56	3.3	

From experiments it is estimated (log-base = 2)

$$1 < H_\infty(X) \leq 1.5 \ !$$

Thinking the Future
Zukunft denken

English Letter Frequencies

The frequencies from this page are generated from around 4.5 billion characters of English text, sourced from Wortschatz. The text files containing the counts can be used with `ngram_score.py` and used for breaking ciphers, see this page for details. If you want to compute the letter frequencies of your own piece of text you can use this page.

English single letter frequencies are as follows (in percent %):

A	8.55	K	0.81	U	2.68
B	1.60	L	4.21	V	1.06
C	3.16	M	2.53	W	1.83
D	3.87	N	7.17	X	0.19
E	12.10	O	7.47	Y	1.72
F	2.18	P	2.07	Z	0.11
G	2.09	Q	0.10		
H	4.96	R	6.33		
I	7.33	S	6.73		
J	0.22	T	8.94		

Digrams. The top 30 are the following (in percent %):

TH	2.71	EN	1.13	NG	0.89
HE	2.33	AT	1.12	AL	0.88
IN	2.03	ED	1.08	IT	0.88
ER	1.78	ND	1.07	AS	0.87
AN	1.61	TO	1.07	IS	0.86
RE	1.41	OR	1.06	HA	0.83
ES	1.32	EA	1.00	ET	0.76
ON	1.32	TI	0.99	SE	0.73
ST	1.25	AR	0.98	OU	0.72
NT	1.17	TE	0.98	OF	0.71

Trigrams. The top 30 are the following (in percent %):

THE	1.81	ERE	0.31	HES	0.24
AND	0.73	TIO	0.31	VER	0.24
ING	0.72	TER	0.30	HIS	0.24
ENT	0.42	EST	0.28	OFT	0.22
ION	0.42	ERS	0.28	ITH	0.21
HER	0.36	ATI	0.26	FTH	0.21
FOR	0.34	HAT	0.26	STH	0.21
THA	0.33	ATE	0.25	OTH	0.21
NTH	0.33	ALL	0.25	RES	0.21
INT	0.32	ETH	0.24	ONT	0.20

Quadgrams. The top 30 are the following (in percent %):

TION	0.31	OTHE	0.16	THEM	0.12
NTHE	0.27	TTHE	0.16	RTHE	0.12
THER	0.24	DTHE	0.15	THEP	0.11
THAT	0.21	INGT	0.15	FROM	0.10
OFTH	0.19	ETHE	0.15	THIS	0.10
FTHE	0.19	SAND	0.14	TING	0.10
THES	0.18	STHE	0.14	THEI	0.10
WITH	0.18	HERE	0.13	NGTH	0.10
INTH	0.17	THEC	0.13	IONS	0.10
ATIO	0.17	MENT	0.12	ANDT	0.10

Quintgrams: The top 30 are the following (in percent %):

OFTHE	0.18	ANDTH	0.07	CTION	0.05
ATION	0.17	NDTHE	0.07	WHICH	0.05
INTHE	0.16	ONTHE	0.07	THESE	0.05
THERE	0.09	EDTHE	0.06	AFTER	0.05
INGTH	0.09	THEIR	0.06	EOFTH	0.05
TOTHE	0.08	TIONA	0.06	ABOUT	0.04
NGTHE	0.08	ORTHE	0.06	ERTHE	0.04
OTHER	0.07	FORTH	0.06	IONAL	0.04
ATTHE	0.07	INGTO	0.06	FIRST	0.04
TIONS	0.07	THECO	0.05	WOULD	0.04