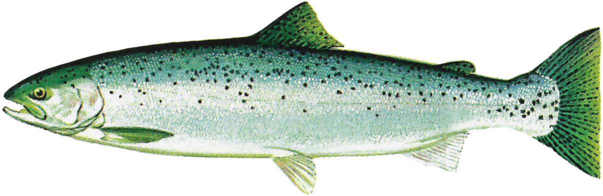
Information Theory
WEDNESDAY

Mathias Winther Madsen
mathias@gmail.com
github.com/mathias-madsen/nasslli2025/



NASSLLI, June 2025

Entropy Rates: Intuition

$X_1, X_2, \ldots, X_{40}$	$\hat{H}_{40}$
SO_SHE_BEGAN_VERY_CAUTIOUSLY:_'BUT_I_DON	4.489
U'VE_HAD!'_'OH,_I'VE_HAD_SUCH_A_CURIOUS_	4.532
F_YOU_PROVIDE_ACCESS_TO_OR_DISTRIBUTE_CO	4.325
'I_DIDN'T_KNOW_THAT_CHESHIRE_CATS_ALWAYS	4.286
LY_ONE_A-PIECE_ALL_ROUND._'BUT_SHE_MUST_	4.369
TS_EYES_WERE_GETTING_EXTREMELY_SMALL_FOR	4.321
ERMS_OF_THE_FULL_PROJECT_GUTENBERG-TM_LI	4.579
W_LARGER,_I_CAN_REACH_THE_KEY;_AND_IF_IT	4.269

$$H = 4.315$$

Entropy Rates: Intuition

$X_1, X_2, \ldots, X_{40}$	$\hat{H}_{40}$
TFIRE_GRTTUOAEUANPR_KAOO_SA'T_H_TW!O_OI_	4.265
SAIU_R,ETDMEI!_SNATLA_'ADESHOMRT_AO___DE	4.185
ERAEAKO,Y_THSEI'OP_ETS__ST_'E_YHCSOSWEON	4.233
___ETEF_OOHRCNEVSNGSLOE_CTOTSINIB_L_ILID	4.190
OEESGE_EUGNA_ES_AUREO_AOTTD_TIARRHAABTTF	4.081
JOE_OAFYNIFTOAOOE__SOFNNATDHOOEETW_IOQPR	4.464
TX_IOWALHALEY_EF_OI_LDDO*_KGHEU_O_FTHEST	4.492
A_TSTSS_WT_LDNER,_SDSUR_U_NTTHEOSRAREDHO	4.113

$$H = 4.315$$

Entropy Rates: Definition

Definition

The **entropy rate** of a random sequence $X_1, X_2, X_3, \ldots$ is

$$\lim_{n \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{n}$$

whenever this limit exists.

Entropy Rates: Examples

Fixed-Length Repetitions

Repeatedly pick a letter at random and print it three times:

`LLL EEE HHH QQQ MMM QQQ OOO TTT EEE YYY XXX GGG ...`

Indefinite Repetition

Pick a letter at random and print it forever:

`WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW ...`

Geometric-Length Repetitions

Repatedly print a (new) random letter as long as a fair coin comes up heads:

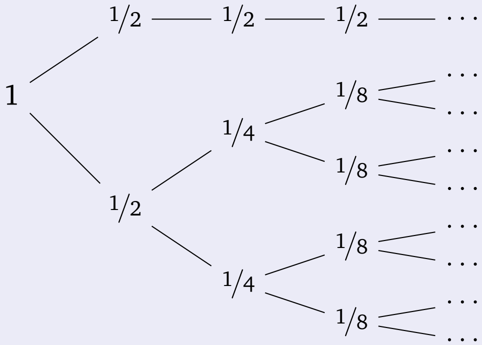
`SSS P MMMMM D HHH K Z T D U C AAA I D TTT Y HHHH ...`

Entropy Rates: Examples

A Non-Stationary Process

2 00 2222 11000111 2222222222222222 ...

A Non-Ergodic Process



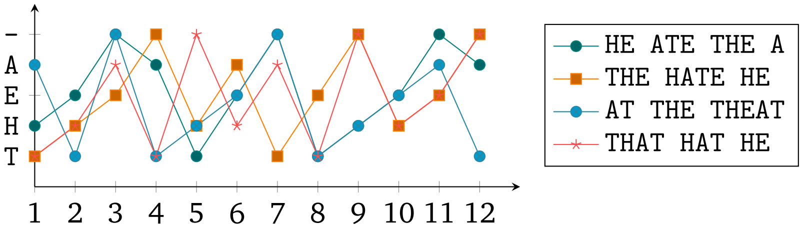
Random Processes: Definition

Definition

A **discrete-time random process** is a countably infinite collection of random variables

$$X_1, X_2, X_3, X_4, \ldots$$

A random process is thus a distribution over **sample paths** $x_1, x_2, x_3, \ldots$.



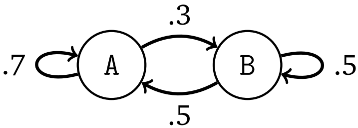
Markov Chains

Definition

A random process P is a **Markov chain** if

$$P(X_{n+1} \,|\, X_1, X_2, \ldots, X_n) \;=\; P(X_{n+1} \,|\, X_n)$$

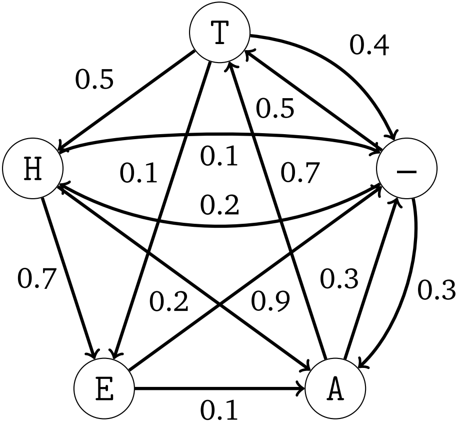
for all n. We call $P(X_{n+1} \,|\, X_n)$ its **transition probabilities**.



$$\begin{array}{cccc} \mathtt{A} & \mathtt{B} & & \\ \downarrow & \downarrow & & \\ .7 & .5 & \to & \mathtt{A} \\ .3 & .5 & \to & \mathtt{B} \end{array}$$

Markov: "On Trials Connected in a Chain of Unobservable Events," *Bulletin of the Imperial Saint Petersburg Academy of Sciences*, 1912.

Markov Chains




```
T_ATE_T_HE_TE_THE_THE_THAT_T_TE_
ATHE_AT_ATHE_T_ATHE_TE_ATH_TH_A_
A_THE_THE_THATEA_THE_HE_A_T_ ...
```

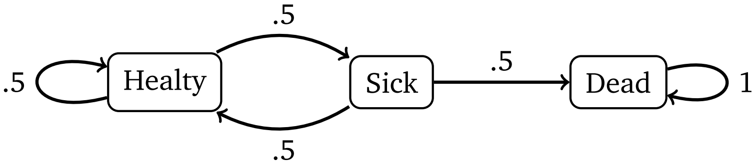
Steady-State Distributions

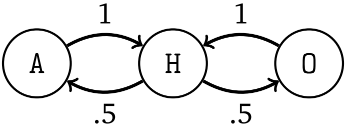
Definition

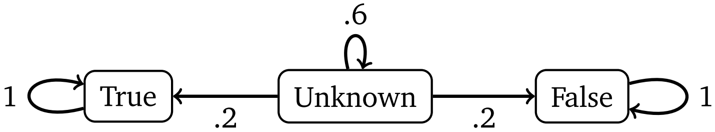
The marginal distribution of X_t is a **steady-state distribution** with respect to the transition probabilities $P(X_{t+1} \mid X_t)$ if X_{t+1} follows the same marginal distribution as X_t.

$$\mathbf{v} = \begin{pmatrix} .3 \\ .4 \\ .3 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} .1 & .3 & .5 \\ .7 & .4 & .1 \\ .2 & .3 & .4 \end{pmatrix}, \quad \mathbf{Tv} = \begin{pmatrix} .3 \\ .4 \\ .3 \end{pmatrix}$$

Steady-State Distributions







Markov Chains

Theorem

In a finite-state Markov chain, the relative visiting frequencies converges to a steady-state distribution.

1		1			1			1
	1		1			1		
				1			1	

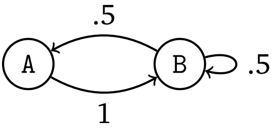
1/1	1/2	1/3	1/4	1/5	2/6	2/7	2/8	$\cdots$
0	1/2	1/3	2/4	2/5	2/6	3/7	3/8	$\cdots$
0	0	0	0	1/5	1/6	1/7	2/8	$\cdots$

Markov Chains

Corollary

If a Markov chain has a unique steady-state distribution P^*, its entropy rate is the average of conditional transition entropies,

$$H = \sum_x P^*(X = x)\, H(X_{t+1} \mid X_t = x)$$



x	A	B
$P^*(X = x)$	1/3	2/3
$H(X_{t+t} \mid X_t = x)$	0	1

$$H = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}.$$

Markov Chains

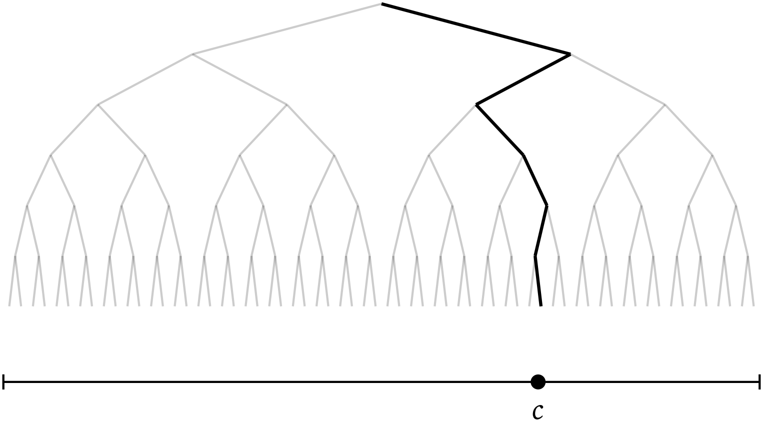
Exercise

Compute the entropy rates of a Markov chain with the transition matrix

$$1.\ \ T = \begin{pmatrix} .6 & .9 \\ .4 & .1 \end{pmatrix} \qquad 2.\ \ T = \begin{pmatrix} .7 & .2 \\ .3 & .8 \end{pmatrix}$$

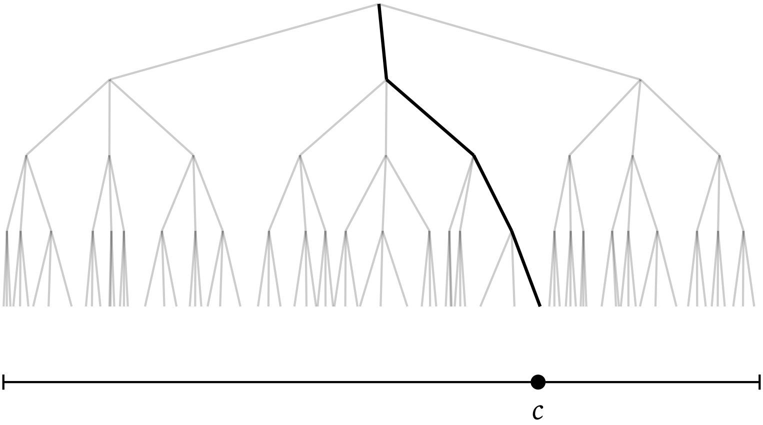
x	.1	.2	.3	.4	.5	.6	.7	.8	.9
$-\log_2 x$	3.3	2.3	1.7	1.3	1.0	0.7	0.5	0.3	0.2

Point Selection



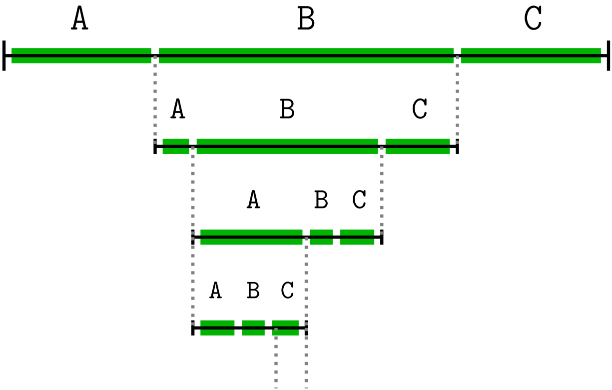
$$c \;=\; \frac{s_1}{2} + \frac{s_2}{4} + \frac{s_3}{8} + \frac{s_4}{16} + \cdots$$

Point Selection



$$c \;=\; \sum_{x<1} p(x) \,+\, p(1) \sum_{x<2} p(x \mid 1) \,+\, p(1,2) \sum_{x<2} p(x \mid 1,2) \,+\, \cdots$$

Point Selection



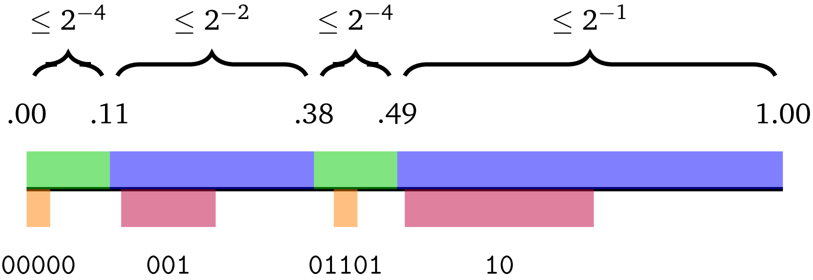
Shannon-Fano-Elias Coding

Shannon-Fano-Elias Code

1. Compute lengths $w_i = \lceil -\log p(x_i) \rceil + 1$
2. Compute cumulatives $c(x) = \sum_{x<x_i} p(x)$
3. Pick nearest binary fraction $b_i/2^{w_i} \geq c(x_i)$

Exercises

1. (.4, .6)
2. (.9, .1)
3. (.3, .4, .3)



Fano: "The Transmission of Information." Technical Report 65, Research Laboratory of Electronics, MIT (1949).

Arithmetic Coding

Cumulative Probability Below $x_1, x_2, \ldots, x_n$

1. Initialize

$$\begin{aligned} c_0 &\leftarrow 0 \\ s_0 &\leftarrow 1 \end{aligned}$$

2. For x_i in $x_1, x_2, \ldots, x_n$:

$$\begin{aligned} c_i &\leftarrow c_{i-1} + s_{i-1} \sum_{x < x_i} p(x \mid x_1, x_2, \ldots, x_{i-1}) \\ s_i &\leftarrow s_{i-1}\, p(x_i \mid x_1, x_2, \ldots, x_{i-1}) \end{aligned}$$

3. Return c_n