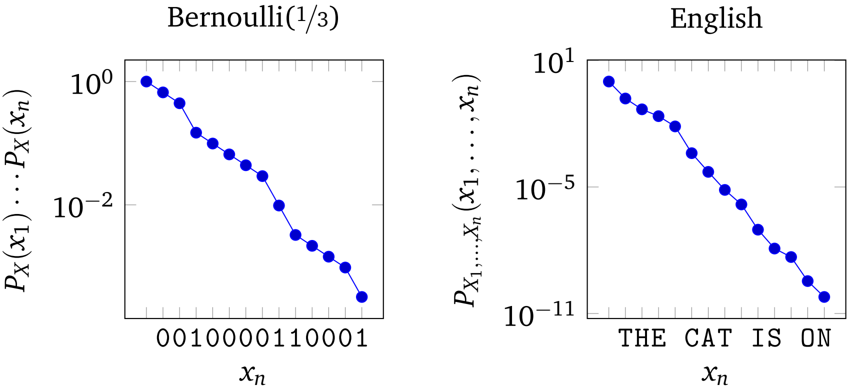
Random Processes and Entropy Rates

Mathias Winther Madsen
mathias.winther@gmail.com

Institute for Logic, Language, and Computation
University of Amsterdam

13 November 2015

Entropy Rates: Intution



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 ...`

Geometric-Length Repetitions

Repatedly print a random letter $k \sim$ Geometric($1/2$) times:

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

Indefinite Repetition

Pick a letter at random and print it forever:

`AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA ...`

Entropy Rates: Examples

A Uniform, Memoryless Process over $\mathcal{X} = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$

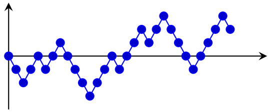
B A C A D A B B D C B B A A D C A C B B A B B D A C B D B B ...

A General Memoryless (i.i.d.) Process

I T T T S S T L C T E C _ E F A I R N P E I A I _ S A R H _ F M ...

Random Walk from $X_1 = 0$

$0, -1, -2, -1, 0, -1, 0, 1, 0, -1, \ldots$



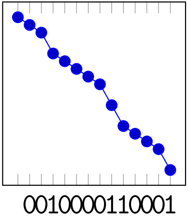
$X_n \sim \text{Uniform}\{1, 2, \ldots, 2^n\}$

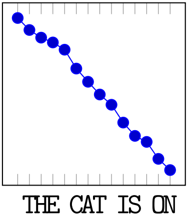
$1, 1, 3, 6, 11, 26, 58, 70, 185, 435, 467, 909, 2804, 5262, \ldots$

Entropy Rates: Here Be Dragons?

Shannon's Source Coding Theorem

In a sequence of i.d.d. samples, the average surprisal converges to the entropy (by the weak law of large numbers).

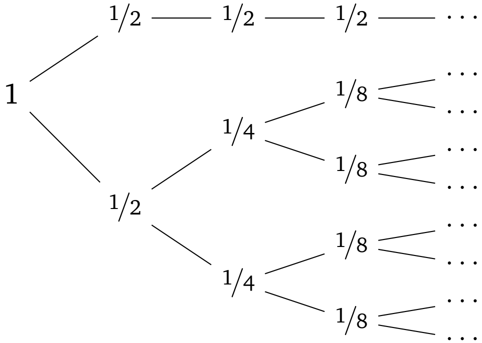




Theorem . . . ?

In a sequence of dependent samples, the average surprisal converges to the entropy rate . . . ?

Entropy Rates: Here Be Dragons?



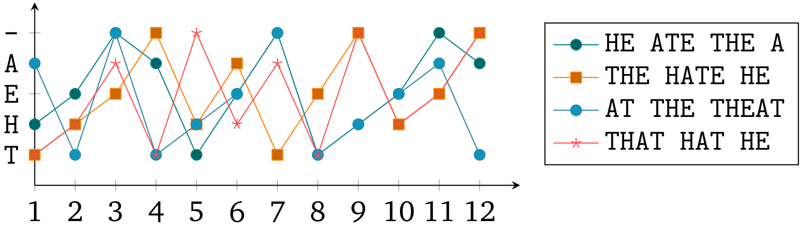
Random Processes: Definition

Definition

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

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

with values in some discrete set $\mathcal{X}$. A random process is thus a distribution over the set of **sample paths** $x_1, x_2, x_3, \ldots$.



Random Processes: Finite Projections

The Daniell-Kolmogorov Extension Theorem

If two random processes assign the same probabilities to all initial-segment events of the form

$$X_1 \in A_1,\ X_2 \in A_2,\ \ldots,\ X_n \in A_n,$$

then they are identical.

P. J. Daniell: "Integrals in An Infinite Number of Dimensions" (*Annals of Mathematics*, Vol. 20(4), 1919).

A. Kolmogorov: *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, 1933), Chapters 2.2 and 3.4.

Markov Chains: Definition

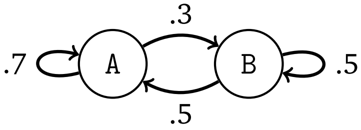
Definition

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

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

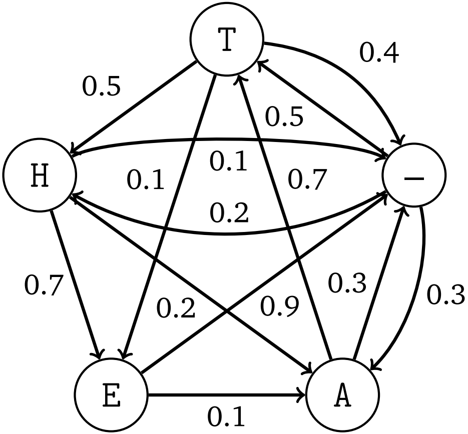
for all n. We call $P(X_{n+1} \mid X_n)$ its **transition probabilities**.

We often assume constant transition probabilities.



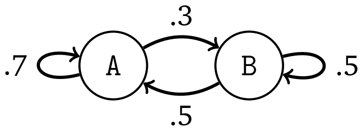
A	B		
$\downarrow$	$\downarrow$		
.7	.5	$\rightarrow$	A
.3	.5	$\rightarrow$	B

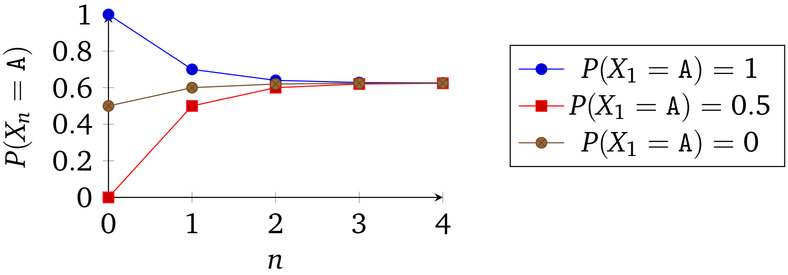
Markov Chains: Modeling



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_ ...

Markov Chains: Stationarity





Markov Chains: Stationarity

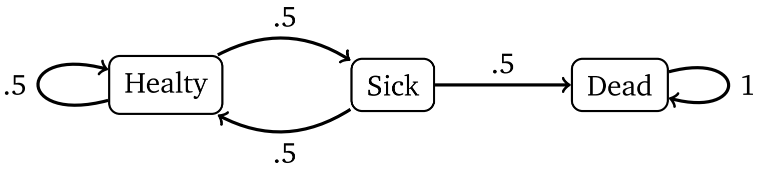
Definition

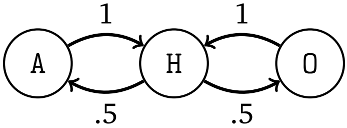
A random process P is **stationary** if

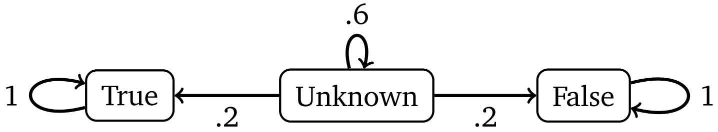
$$P(X_1 = x_1, \ldots, X_n = x_n) \;=\; P(X_2 = x_1, \ldots, X_{n+1} = x_n)$$

for all n and all value vectors $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$.

Markov Chains: Stationarity







Time-Averages

Definition

The nth **time-average** of a (measurable) function $f : \mathcal{X}^{\mathbb{N}} \to \mathbb{R}$ on the sample path $x = x_1, x_2, x_3, \ldots$ is

$$A_n f(x) \;=\; \frac{f(x_1, x_2, \ldots) + f(x_2, x_3, \ldots) + \cdots + f(x_n, x_{n+1}, \ldots)}{n}.$$

The **limiting time-average** on x is $\lim_{n\to\infty} A_n f(x)$.

Main example:

$$f(x_1, x_2, x_3, \ldots) \;=\; \begin{cases} 1 & (x_1 \in A) \\ 0 & (x_1 \notin A) \end{cases}$$

Convergence: Existence

The "Ergodic Theorem"

If a random process is stationary, then its time-averages converge with probability 1.

J. von Neumann: "Proof of the Quasi-ergodic Hypothesis" (*Proceedings of the Natural Academy of Sciences of the USA*, Vol. 18(1), 1932).

G. D. Birkhoff: "Proof of the ergodic theorem" (*Proceedings of the Natural Academy of Sciences of the USA*, Vol. 17(12), 1931).

Time-Invariance

Definition

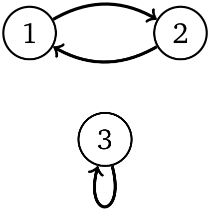
A set B of sample paths is called **time-invariant** if

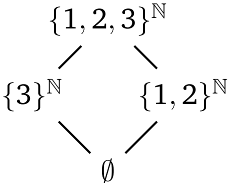
$$(x_1, x_2, x_3, \ldots) \in B \qquad \Longrightarrow \qquad (x_2, x_3, x_4, \ldots) \in B$$

Time-invariant predicates of x:

1. The sample path x never visits the set $A \subseteq \mathcal{X}$.
2. The sample path x visits the set $A \subseteq \mathcal{X}$ infinitely often.
3. The sample path x is constant, $x_1 = x_2 = x_3 = \cdots$.
4. The sample path x eventually enters a trapping set $A \subseteq \mathcal{X}$ and never leaves.
5. The sample path x passes through $A \subseteq \mathcal{X}$ with a relative frequency that converges to f^*.

Time-Invariance





x	$P(X = x)$
$1, 2, 1, 2, 1, 2, \ldots$	$1/3$
$2, 1, 2, 1, 2, 1, \ldots$	$1/3$
$3, 3, 3, 3, 3, 3, \ldots$	$1/3$

Convergence: Uniqueness

Definition

A random process P is be **ergodic** if it assigns probability 0 or 1 to all time-invariant sets.

Uniqueness of Averages

Under an ergodic process, limiting time-averages are almost constant (i.e., take the same fixed value with probability 1).

(*Proof*: From the cumulative distribution of $\lim_n A_n f(X)$.)

Time-Averaged Surprisal

The Shannon-McMillan-Breiman Theorem

On a sample path drawn from a stationary and ergodic random process, the average surprisal converges to the entropy rate with probability 1.

B. McMillan: "The basic theorems of information theory" (*Annals of Mathematical Statics*, Vol. 24, 1953).

L. Breiman: "The individual ergodic theorem of information theory" (*Annals of Mathematical Statics*, Vol. 28, 1957).

Time-Averaged Surprisal

Half-Deterministic: $\frac{1}{2}\text{Bernoulli}(0) + \frac{1}{2}\text{Bernoulli}(1/2)$

$$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots$$
$$1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, \ldots$$

A Stationary Markov Chain

`ATHE_AT_ATHE_T_ATHE_TE_ATH_TH_A_A_THE ...`

Random Walk from $X_1 = 0$

$$0, 1, 2, 3, 4, 3, 2, 3, 2, 1, 2, 1, 0, -1, -2, -1, 0, -1, -2, \ldots$$

Non-Ergodic Processes

Definition

Two distributions P_1 and P_2 are **mutually singular** if they have disjoint supports.

Partitioning

Two stationary and ergodic processes P_1^* and P_2^* are either identical or mutually singular.

(*Proof*: By projection to a finite-dimensional event.)

Non-Ergodic Processes

Definition

A distribution P is **absolutely continuous** with respect to a reference distribution P^* if

$$P^*(B) = 0 \qquad \Longrightarrow \qquad P(B) = 0$$

Attractor Processes

If a random process P is absolutely continuous with respect to a stationary and ergodic process P^*, then their limiting time-averages coincide.

(*Proof*: $P^*(f^*) = 1$, so $P(f^*) = 1$ by absolute continuity.)

Non-Ergodic Processes

Ups and Downs

Repeatedly print $k \sim$ Geometric(1/2) left-parentheses and immediately after, k right-parentheses:

$$()((()))((()))(())(())((()))(()) \ldots$$

Beta Urn

Draw a marble from an urn with 5 blue and 5 red marbles; add an extra marble of the same color to the urn; repeat:

R R R B R R B R R R B R R B B B R R R B B R B R R R R R B B . . .

$X_n \sim \text{Uniform}\{1, 2, \ldots, 2^n\}$

$$1, 1, 3, 6, 11, 26, 58, 70, 185, 435, 467, 909, 2804, 5262, \ldots$$