Algorithmic Complexity
[Zurek(1989)]

Marek Pikulski

Proseminar *The Role of Information in Statistical Mechanics*,
Prof. Renner, ETH Zürich

May 25, 2009

Bibliography

W. H. Zurek.
Algorithmic randomness and physical entropy.
Phys. Rev. A, 40:4731, 1989.

L. Szilard.
Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen.
Zeitschrift für Physik A, 53:840–856, 1928.

R. Landauer.
Irreversibility and heat generation in the computing process.
IBM Journal of Research and Development, 5:183–191, 1961.

C. H. Bennett.
Logical reversibility of computation.
IBM Journal of Research and Development, 17:525–532, 1973.

Outline

Reminder of statistical entropy and motivation

Entropy in statistical mechanics

Gibbs entropy[a]

$$H = -k_B \sum_i p_i \log p_i \quad \text{or} \quad H = -\sum_i p_i \log_2 p_i$$

Boltzmann formula (for W equiprobable states): $H = \log_2 W$
von Neumann entropy: $H = -Tr\rho \ln \rho$

[a] formula depends on "choice of units"

Entropy in information theory

Shannon entropy (characterized by few mathematical properties)

$$H = -\sum_i p_i \log_2 p_i$$

Motivation

Statistical entropy requires specification of an ensemble and a probability distribution!

Question: What is the entropy of a single, definite microstate?

- Use algorithmic randomness to define it.
- $\longrightarrow$ Study its properties.
- Try to unify statistical and algorithmic entropy.

Outline

The configuration of a system can be specified by a binary string.
$\longrightarrow$ What is the (algorithmic) complexity of such a string?

Example

- 010101010101 is algorithmically simple
- 100010111000 is algorithmically random
- However, the leading bits of a binary representation of $\sqrt{(}2)$ are rather simple.

The configuration of a system can be specified by a binary string.
$\rightarrow$ What is the (algorithmic) complexity of such a string?

Example

- 010101010101 is algorithmically simple
- 100010111000 is algorithmically random
- However, the leading bits of a binary representation of $\sqrt(2)$ are rather simple.

Definition (Algorithmic randomness)

Let s be a binary string and U a universal Turing machine. The *algorithmic randomness of s*, $K(s)$, is the length $|s^*|$ of the shortest program s^*, executable by U, that has the following properties:

i) s^* terminates (after finite time)

ii) s^* is self-delimiting (contains its length)

iii) after the execution of s^* the output band contains nothing but s

Definition

A computer/Turing machine U is said to be *universal* iff for every other computer C there exists a prefix τ_C so that $U(\tau_C p) = C(p)$ for all programs p which can be executed by C.

First remarks concerning algorithmic randomness

- Most binary strings s are *algorithmically random* ($K(s) \approx |s|$).
- Binary representations of *random* integers i:

$$K(i) \quad \approx \quad \log_2 i + \mathcal{O}(1)$$

Example (Boltzmann Gas)

- Ideal gas in D-dimensional container of fixed volume V
- N indistinguishable particles at temperature T
- Partition phase space into equally sized, rectangular, paraxial cells of dimensions $\Delta_V = {\Delta_x}^D$ and Δ_p.
- $E = \frac{1}{2} k_B T$ per degree of freedom $\Rightarrow p_i = \sqrt{m k_B T}$

Number of grid vertices:

$$C \approx \left(\frac{V}{\Delta_V} \right) \left(\frac{\sqrt{m k_B T}}{\Delta_p} \right)^D$$

Total number of distinguishable configurations:

$$\Omega = \frac{C^N}{N!}$$

Example (Boltzmann Gas)

- Ideal gas in D-dimensional container of fixed volume V
- N indistinguishable particles at temperature T
- Partition phase space into equally sized, rectangular, paraxial cells of dimensions $\Delta_V = {\Delta_x}^D$ and Δ_p.
- $E = \frac{1}{2} k_B T$ per degree of freedom $\Rightarrow p_i = \sqrt{m k_B T}$

Number of grid vertices:

$$C \approx \left(\frac{V}{\Delta_V}\right) \left(\frac{\sqrt{m k_B T}}{\Delta_p}\right)^D$$

Total number of distinguishable configurations:

$$\Omega = \frac{C^N}{N!}$$

Algorithmic randomness of a typical configuration [1]:

$$K \approx \log_2(\Omega) + \mathcal{O}(1) = N\left(\log_2\left(\frac{V}{\sqrt[N]{N!}\Delta_V}\right) + \frac{D}{2}\log_2\frac{mk_BT}{\Delta_p^2}\right) + \mathcal{O}(1)$$

$$\overset{\text{Stirling, big } N}{\approx} N\left(\log_2\left(\frac{V}{N\Delta_V}\right) + \frac{D}{2}\log_2\frac{mk_BT}{\Delta_p^2}\right) + \mathcal{O}(N) + \mathcal{O}(1)$$

$$\overset{\text{huge } N}{\approx} N\left(\log_2\left(\frac{V}{N\Delta_V}\right) + \frac{D}{2}\log_2\frac{mk_BT}{\Delta_p^2}\right) + \mathcal{O}(1)$$

Consistent with the Sackur-Tetrode equation!

[1] Stirling's formula: $\log N! \approx N\log N - N$

Outline

Algorithmic randomness and statistical entropy

What do we want to do?

We have:

- $\langle K \rangle_{\mathcal{E}}$ (average algorithmic randomness)
- $H(\mathcal{E})$ (Shannon entropy)

How are these two quantities related to each other?

Results from information and coding theory

Coding theory

Efficiently encode symbols $\{s_k\}$, occurring with probabilities p_k, using code words $\{\tilde{s_k}\}$ (composed of an alphabet, here $\{0, 1\}$).
Measure of efficiency: $\mathcal{L} = \langle|\tilde{s_k}|\rangle = \sum_i p_i|\tilde{s_i}|$

Definitions

Unique decodability Mapping "sequence of symbols" $\mapsto$ "sequence of code words" is injective.

Instantaneous code Each symbol can be decoded immediately after reception.

Prefix-free code No code word is prefix to any other code word[a].

[a] implies *instantaneous code*

Example

$s_1 \to 0$, $s_2 \to 1$, $s_3 \to 00$, $s_4 \to 11$ not uniquely decodable

$s_1 \to 0$, $s_2 \to 01$, $s_3 \to 011$, $s_4 \to 111$ uniquely dec. , not instantaneous [a]

$s_1 \to 0$, $s_2 \to 10$, $s_3 \to 110$, $s_4 \to 111$ instantaneous, prefix-free

[a] consider **00111**

Definitions

Unique decodability	Mapping "sequence of symbols" $\mapsto$ "sequence of code words" is injective.
Instantaneous code	Each symbol can be decoded immediately after reception.
Prefix-free code	No code word is prefix to any other code word[a].

[a] implies *instantaneous code*

Example

$s_1 \rightarrow 0$, $s_2 \rightarrow 1$, $s_3 \rightarrow 00$, $s_4 \rightarrow 11$ not uniquely decodable

$s_1 \rightarrow 0$, $s_2 \rightarrow 01$, $s_3 \rightarrow 011$, $s_4 \rightarrow 111$ uniquely dec. , not instantaneous [a]

$s_1 \rightarrow 0$, $s_2 \rightarrow 10$, $s_3 \rightarrow 110$, $s_4 \rightarrow 111$ instantaneous, prefix-free

[a] consider **00111**

Definitions

Unique decodability	Mapping "sequence of symbols" $\mapsto$ "sequence of code words" is injective.
Instantaneous code	Each symbol can be decoded immediately after reception.
Prefix-free code	No code word is prefix to any other code word[a].

[a] implies *instantaneous code*

Example

$s_1 \rightarrow 0$, $s_2 \rightarrow 1$, $s_3 \rightarrow 00$, $s_4 \rightarrow 11$ not uniquely decodable

$s_1 \rightarrow 0$, $s_2 \rightarrow 01$, $s_3 \rightarrow 011$, $s_4 \rightarrow 111$ uniquely dec. , not instantaneous [a]

$s_1 \rightarrow 0$, $s_2 \rightarrow 10$, $s_3 \rightarrow 110$, $s_4 \rightarrow 111$ instantaneous, prefix-free

[a] consider **001**11

Definitions

Unique decodability	Mapping "sequence of symbols" $\mapsto$ "sequence of code words" is injective.
Instantaneous code	Each symbol can be decoded immediately after reception.
Prefix-free code	No code word is prefix to any other code word[a].

[a] implies *instantaneous code*

Example

$s_1 \to 0,\ s_2 \to 1,\ s_3 \to 00,\ s_4 \to 11$ not uniquely decodable

$s_1 \to 0,\ s_2 \to 01,\ s_3 \to 011,\ s_4 \to 111$ uniquely dec. , not instantaneous [a]

$s_1 \to 0,\ s_2 \to 10,\ s_3 \to 110,\ s_4 \to 111$ instantaneous, prefix-free

[a] consider **001**11

Theorem (Kraft's inequality)

Every uniquely decodable code (not necessarily finite number of symbols) that uses an alphabet consisting of q letters satisfies the inequality

$$\sum_i q^{-l_i} \leq 1$$

Conversely, for every given set of code word lengths $\{l_i\}$ which satisfy the preceding inequality there exists a (even prefix-free) code.

Shannon coding

Shannon coding is defined (up to permutations of code words of equal length) by $l_k = \lceil -\log_2 p_k \rceil$ and can be generated as shown in the proof of the converse of Kraft's inequality.

$\tilde{s}_k$ can be generated by knowing only the symbols/probabilities with $l_i \leq l_k$.

Theorem (Shannon's source coding theorem)

For every decipherable (uniquely decodable) code with word lengths l_k which encodes symbols s_k occurring with probabilities p_k:

$$H = -\sum_k p_k \log_2 p_k \quad \leq \quad \sum_k p_k l_k = \mathcal{L}$$

If the code is optimal (minimizes the expected word length $\mathcal{L}$):

$$H \quad \leq \quad \mathcal{L} \quad \leq \quad H + 1 \qquad .$$

Definition

The *algorithmic information content of an ensemble* $\mathcal{E}$ consisting of states s_k with probabilities p_k, $K(\mathcal{E})$, is the length of the shortest program $\mathcal{E}^*$ for a universal Turing machine which is able to enumerate the states s_k—each of them up to a given precision—along with the values of p_k (up to a given precision).

The output of $\mathcal{E}^*$ has to be *weakly sorted*: for every $\delta > 0$ there should exist a finite number of steps N_δ after which $\mathcal{E}^*$ has listed all states with probabilities $p_k > \delta$.

Definition

An ensemble $\mathcal{E}$ is called *thermodynamic* if $K(\mathcal{E}) \lll H(\mathcal{E})$, where H is the statistical entropy.

Theorem (Ensemble averages of algorithmic randomness)

The ensemble average of the algorithmic randomness for an ensemble $\mathcal{E}$, $\langle K \rangle_{\mathcal{E}}$, is bounded from below and from above in terms of the statistical entropy of the ensemble $H(\mathcal{E})$ and the algorithmic information content of the ensemble specification $K(\mathcal{E})$ according to the following inequality.

$$H(\mathcal{E}) \quad \leq \quad \langle K \rangle_{\mathcal{E}} \quad \leq \quad H(\mathcal{E}) + K(\mathcal{E}) + O(1)$$

Proof.

Blackboard. □

Theorem (Shannon's source coding theorem)

For every decipherable (uniquely decodable) code with word lengths l_k which encodes symbols s_k occurring with probabilities p_k:

$$H = -\sum_k p_k \log_2 p_k \quad \leq \quad \sum_k p_k l_k = \mathcal{L}$$

If the code is optimal (minimizes the expected word length $\mathcal{L}$):

$$H \quad \leq \quad \mathcal{L} \quad \leq \quad H + 1 \qquad .$$

Theorem (Ensemble averages of algorithmic randomness)

The ensemble average of the algorithmic randomness for an ensemble $\mathcal{E}$, $\langle K \rangle_{\mathcal{E}}$, is bounded from below and from above in terms of the statistical entropy of the ensemble $H(\mathcal{E})$ and the algorithmic information content of the ensemble specification $K(\mathcal{E})$ according to the following inequality.

$$H(\mathcal{E}) \quad \leq \quad \langle K \rangle_{\mathcal{E}} \quad \leq \quad H(\mathcal{E}) + K(\mathcal{E}) + O(1)$$

Proof.

Blackboard. □

Corollary

If $\mathcal{E}$ is a thermodynamic ensemble ($K(\mathcal{E}) \lll H(\mathcal{E})$), the preceding theorem yields

$$\langle K \rangle_{\mathcal{E}} \approx H(\mathcal{E})$$

Outline

Physical entropy, Motivation

Consider a thermodynamic engine and an IGUS (information gathering and using system) that is able to

1. perform measurements
2. perform computations using the measurement results
3. optimize the operation of the engine

This model is inspired by Szilard's engine [Szilard(1928)].

Definition

Physical entropy is the sum of the missing information and the size of the most concise record containing the data d known about a system:

$$\mathcal{S}_d = H_d + K(d)$$

Where, for a system which can be found in states $\{s_k\}$ with respective probabilities $\{p_k\}$, $H(d)$ is given by the conditional Shannon entropy:

$$H_d = -\sum_k p_{k|d} \log_2 p_{k|d}$$

Definition (*Physical entropy*)

sum of missing information and complexity of known data d:

$$\mathcal{S}_d = H_d + K(d), \quad H_d = -\sum_k p_{k|d} \log_2 p_{k|d}$$

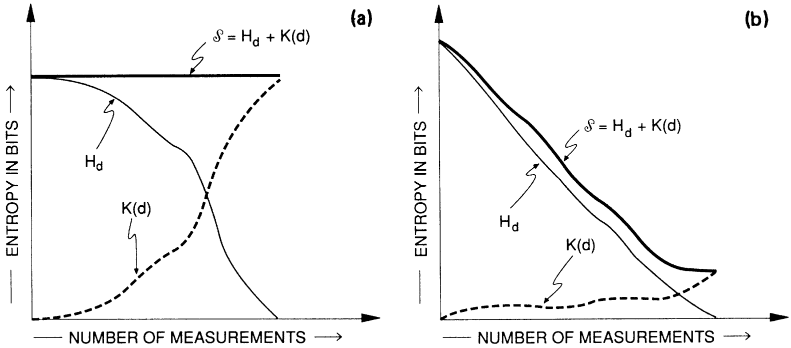

Figure: [Zurek(1989)], fig. 2

Extraction of work by an IGUS

Show that IGUS cannot extract any work in average.

Calculate average value of $\mathcal{S}_d$ for an ensemble $\mathcal{E}$, applying the preceding theorem to the ensemble '$\mathcal{E}$ restricted by fixing d' ($\mathcal{E}|d$):

$$\langle \mathcal{S}_d \rangle_d = \sum_d p_d \left(H_d + K(d)\right)$$

$$\geq \sum_d p_d \langle K \rangle_{\mathcal{E}|d} = \sum_{k,d} p_d \, p_{k|d} \, K(s_k) = \sum_k p_k K(s_k) = \langle K \rangle_{\mathcal{E}}$$

And:

$$\langle \mathcal{S}_d \rangle_d = \sum_d p_d \left(H_d + K(d)\right) \leq \sum_d p_d \left(\langle K \rangle_{\mathcal{E}|d} + K(d)\right)$$

$$= \sum_{k,d} p_d \, p_{k|d} \, K(s_k) + \sum_d p_d \, K(d) = \langle K \rangle_{\mathcal{E}} + \langle K(d) \rangle_d$$

According to [Zurek(1989)], the first inequality is in fact approximately an equality.

"The demon's version of thermodynamics"

Try to justify the term "entropy" in "physical entropy".

Thermodynamics

Thermodynamic entropy S defined by:

$$\begin{aligned} dU &= -\delta W + \delta Q \\ \delta Q &= T\,dS \end{aligned}$$

(δW work done by the system, δQ heat transferred to the system, U internal energy, T temperature)

Show that physical entropy $\mathcal{S}$ satisfies

$$\Delta W = T\,\Delta\mathcal{S} \quad .$$

(Assume $dU = 0$ and $T = const.$ for simplicity)

"The demon's version of thermodynamics"

Try to justify the term "entropy" in "physical entropy".

Thermodynamics

Thermodynamic entropy S defined by:

$$\begin{aligned} dU &= -\delta W + \delta Q \\ \delta Q &= T\, dS \end{aligned}$$

(δW work done by the system, δQ heat transferred to the system, U internal energy, T temperature)

Show that physical entropy $\mathcal{S}$ satisfies

$$\Delta W = T\, \Delta \mathcal{S} \quad .$$

(Assume $dU = 0$ and $T = const.$ for simplicity)

- Consider a transition of the system from state s_i to state s_f in the presence of an demon-type observer (IGUS)—think of Szilard's engine.
- Assume IGUS operates at temperature T.
- Assume that the demon always keeps its memory record about the state of the system (r) up-to-date.
- Initially, $r = r_i$. After the transition, $r = r_f$.

Energy the demon can gain due to the change of (statistical) entropy:

$$\Delta W^+ = T(H_f - H_i)$$

- Consider a transition of the system from state s_i to state s_f in the presence of an demon-type observer (IGUS)—think of Szilard's engine.
- Assume IGUS operates at temperature T.
- Assume that the demon always keeps its memory record about the state of the system (r) up-to-date.
- Initially, $r = r_i$. After the transition, $r = r_f$.

Energy the demon can gain due to the change of (statistical) entropy:

$$\Delta W^{+} = T(H_f - H_i)$$

Landauer's principle [Landauer(1961)]

The erasure of 1 bit of information requires an energy of at least $k_B T \ln 2$. This is equal to T in the presented treatment.

What is ΔW^{-}? (What is the minimum number of bits that have to be erased when updating r_i with r_f?)

The substitution of r with r^* (assuming r^* is known) can be achieved by reversible computation (see [Bennett(1973)]):

$$\left(r, \quad r^*\right) \stackrel{\text{rev.}}{\rightarrow} \left(r, \quad r\right) \stackrel{\text{rev.}}{\rightarrow} \left(r, \quad 0\right) \underset{1^{st}\text{ calc., rev.}}{\xrightarrow{\text{backtrack}}} \left(r^*, \quad 0\right)$$

Assuming mapping $r_i \mapsto r_f$ (operating procedure of the engine) is "hard-coded" into IGUS:

$$K(r_i, r_f) = K(r_i)$$
$$K(r_i, r_f) = K(r_f) + K(r_i|r_f^*) \quad . \qquad \text{(note: } r_f^* \text{ is self-delimiting)}$$

- $r_i \rightarrow r_f^* \, r_{i|f*}^*$ reversible ($r_{i|f*}^*$: min. program to calc. r_i given r_f^*).
- $r_f^* \, r_{i|f*}^* \rightarrow r_f \, \underbrace{r_{i|f*}^*}$ reversible

  needs to be erased

  has minimal length
- Minimal thermodyn. cost of memory update:
  $\Delta W^- = |r_{i|f*}^*| = K(r_i|r_f^*) = K(r_i) - K(r_f)$

Joint information satisfies:

$$K(s,t) \leq K(s) + K(t) + \mathcal{O}(1)$$

Definition of conditional information:

$$K(s,t) = K(t) + K(s|(t, K(t))) + \mathcal{O}(1)$$

The substitution of r with r^* (assuming r^* is known) can be achieved by reversible computation (see [Bennett(1973)]):

$$\begin{pmatrix} r, & r^* \end{pmatrix} \stackrel{\text{rev.}}{\longrightarrow} \begin{pmatrix} r, & r \end{pmatrix} \stackrel{\text{rev.}}{\longrightarrow} \begin{pmatrix} r, & 0 \end{pmatrix} \xrightarrow[1^{st}\text{ calc., rev.}]{\text{backtrack}} \begin{pmatrix} r^*, & 0 \end{pmatrix}$$

Assuming mapping $r_i \mapsto r_f$ (operating procedure of the engine) is "hard-coded" into IGUS:

$$\begin{aligned} K(r_i, r_f) &= K(r_i) \\ K(r_i, r_f) &= K(r_f) + K(r_i | r_f^*) \quad . \end{aligned}$$

(note: r_f^* is self-delimiting)

- $r_i \longrightarrow r_f^* \, r_{i|f*}^*$ reversible ($r_{i|f*}^*$: min. program to calc. r_i given r_f^*).
- $r_f^* \, r_{i|f*}^* \longrightarrow r_f \underbrace{r_{i|f*}^*}$ reversible

  needs to be erased

  has minimal length
- Minimal thermodyn. cost of memory update:
  $\Delta W^- = |r_{i|f*}^*| = K(r_i | r_f^*) = K(r_i) - K(r_f)$

The substitution of r with r^* (assuming r^* is known) can be achieved by reversible computation (see [Bennett(1973)]):

$$\left(r, \quad r^*\right) \overset{\text{rev.}}{\rightarrow} \left(r, \quad r\right) \overset{\text{rev.}}{\rightarrow} \left(r, \quad 0\right) \underset{1^{st}\text{ calc., rev.}}{\overset{\text{backtrack}}{\longrightarrow}} \left(r^*, \quad 0\right)$$

Assuming mapping $r_i \mapsto r_f$ (operating procedure of the engine) is "hard-coded" into IGUS:

$$\begin{aligned} K(r_i, r_f) &= K(r_i) \\ K(r_i, r_f) &= K(r_f) + K(r_i | r_f^*) \quad . \qquad \text{(note: } r_f^* \text{ is self-delimiting)} \end{aligned}$$

- $r_i \rightarrow r_f^* \; r_{i|f*}^*$ reversible ($r_{i|f*}^*$: min. program to calc. r_i given r_f^*).
- $r_f^* \; r_{i|f*}^* \rightarrow r_f \; \underbrace{r_{i|f*}^*}$ reversible

  needs to be erased

  has minimal length

- Minimal thermodyn. cost of memory update:
  $\Delta W^- = |r_{i|f*}^*| = K(r_i | r_f^*) = K(r_i) - K(r_f)$

The substitution of r with r^* (assuming r^* is known) can be achieved by reversible computation (see [Bennett(1973)]):

$$\left(r, \quad r^*\right) \stackrel{\text{rev.}}{\rightarrow} \left(r, \quad r\right) \stackrel{\text{rev.}}{\rightarrow} \left(r, \quad 0\right) \underset{1^{st}\text{ calc., rev.}}{\xrightarrow{\text{backtrack}}} \left(r^*, \quad 0\right)$$

Assuming mapping $r_i \mapsto r_f$ (operating procedure of the engine) is "hard-coded" into IGUS:

$$\begin{aligned} K(r_i, r_f) &= K(r_i) \\ K(r_i, r_f) &= K(r_f) + K(r_i|r_f^*) \quad . \qquad \text{(note: } r_f^* \text{ is self-delimiting)} \end{aligned}$$

- $r_i \longrightarrow r_f^*\, r_{i|f^*}^*$ reversible ($r_{i|f^*}^*$: min. program to calc. r_i given r_f^*).
- $r_f^*\, r_{i|f^*}^* \longrightarrow r_f\, \underbrace{r_{i|f^*}^*}$ reversible

  needs to be erased

  has minimal length
- Minimal thermodyn. cost of memory update:
  $\Delta W^- = |r_{i|f^*}^*| = K(r_i|r_f^*) = K(r_i) - K(r_f)$

Theorem

The maximal work gained by an engine coupled with a computerized demon, which can perform measurements and control the operation of the engine is no more than

$$\Delta W = \Delta W^{+} - \Delta W^{-} = T(H_f - H_i - K(r_i) + K(r_f)) = T(\mathcal{S}_f - \mathcal{S}_i)$$

Summary

- Boltzmann-Gibbs-Shannon entropy (statistical entropy)
  - objective for a given ensemble
  - requires probability distribution
  - relatively easy to calculate
  - successful/proven in most applications
  - limit of physical entropy in case of full ignorance
- Algorithmic entropy (algorithmic randomness)
  - objective
  - defined for a single, definite microstate
  - difficult to calculate, relatively easy to estimate
  - limit of physical entropy in case of complete knowledge
- Physical entropy
  - observer-dependent
  - allows formulation of thermodynamics in the presence of a demon-type observer
  - enables to monitor efficiency of a demon after each single cycle of the engine
  - compatible with statistical and algorithmic entropy!

Summary

- Boltzmann-Gibbs-Shannon entropy (statistical entropy)
    - objective for a given ensemble
    - requires probability distribution
    - relatively easy to calculate
    - successful/proven in most applications
    - limit of physical entropy in case of full ignorance
- Algorithmic entropy (algorithmic randomness)
    - objective
    - defined for a single, definite microstate
    - difficult to calculate, relatively easy to estimate
    - limit of physical entropy in case of complete knowledge
- Physical entropy
    - observer-dependent
    - allows formulation of thermodynamics in the presence of a demon-type observer
    - enables to monitor efficiency of a demon after each single cycle of the engine
    - compatible with statistical and algorithmic entropy!

Summary

- Boltzmann-Gibbs-Shannon entropy (statistical entropy)
  - objective for a given ensemble
  - requires probability distribution
  - relatively easy to calculate
  - successful/proven in most applications
  - limit of physical entropy in case of full ignorance
- Algorithmic entropy (algorithmic randomness)
  - objective
  - defined for a single, definite microstate
  - difficult to calculate, relatively easy to estimate
  - limit of physical entropy in case of complete knowledge
- Physical entropy
  - observer-dependent
  - allows formulation of thermodynamics in the presence of a demon-type observer
  - enables to monitor efficiency of a demon after each single cycle of the engine
  - compatible with statistical and algorithmic entropy!

Summary

- Boltzmann-Gibbs-Shannon entropy (statistical entropy)
  - objective for a given ensemble
  - requires probability distribution
  - relatively easy to calculate
  - successful/proven in most applications
  - limit of physical entropy in case of full ignorance
- Algorithmic entropy (algorithmic randomness)
  - objective
  - defined for a single, definite microstate
  - difficult to calculate, relatively easy to estimate
  - limit of physical entropy in case of complete knowledge
- Physical entropy
  - observer-dependent
  - allows formulation of thermodynamics in the presence of a demon-type observer
  - enables to monitor efficiency of a demon after each single cycle of the engine
  - compatible with statistical and algorithmic entropy!

Summary

- Boltzmann-Gibbs-Shannon entropy (statistical entropy)
  - objective for a given ensemble
  - requires probability distribution
  - relatively easy to calculate
  - successful/proven in most applications
  - limit of physical entropy in case of full ignorance
- Algorithmic entropy (algorithmic randomness)
  - objective
  - defined for a single, definite microstate
  - difficult to calculate, relatively easy to estimate
  - limit of physical entropy in case of complete knowledge
- Physical entropy
  - observer-dependent
  - allows formulation of thermodynamics in the presence of a demon-type observer
  - enables to monitor efficiency of a demon after each single cycle of the engine
  - compatible with statistical and algorithmic entropy!