Algorithmic Information Theory and Foundations of probability

Alexander Shen (LIF, Marseille, on leave from IITP, Moscow)

September 23–25, 2009, RP-09

Acknowledgements and apology

Natural Science: a simplistic view

Theoretician

Experimenter

Natural Science: a simplistic view

Theoretician	Experimenter
predictions (theory)	observations (data)

Natural Science: a simplistic view

Theoretician	Experimenter
predictions (theory)	observations (data)

match or not

Probability theory

predictions | observations

Probability theory

predictions	observations
Probability distribution	outcome

Probability theory

predictions	observations
Probability distribution	outcome
fair coin (uniform on $\mathbb{B}^{1000}$)	010101…01

Probability theory

predictions	observations
Probability distribution	outcome
fair coin (uniform on $\mathbb{B}^{1000}$)	010101...01

do they match?

Probability theory

predictions	observations
Probability distribution	outcome
fair coin (uniform on $\mathbb{B}^{1000}$)	$010101\ldots01$

do they match?

No, since the probability of outcome $0101\ldots01$ is negligible (2^{-1000}).

Probability theory

predictions	observations
Probability distribution	outcome
fair coin (uniform on $\mathbb{B}^{1000}$)	010101…01

do they match?

No, since the probability of outcome $0101\ldots01$ is negligible (2^{-1000}).

OK, but this is the case for *any* outcome!

“Shuffle machine” paradox

“Shuffle machine” paradox

- A casino uses fresh well-shuffled deck of cards of each game

"Shuffle machine" paradox

- A casino uses fresh well-shuffled deck of cards of each game
- Outsourcing shuffling: shrink-wrapped well shuffled decks

"Shuffle machine" paradox

- A casino uses fresh well-shuffled deck of cards of each game
- Outsourcing shuffling: shrink-wrapped well shuffled decks
- Shuffling factory: quality control that blocks decks that are not well shuffled.

"Shuffle machine" paradox

- A casino uses fresh well-shuffled deck of cards of each game
- Outsourcing shuffling: shrink-wrapped well shuffled decks
- Shuffling factory: quality control that blocks decks that are not well shuffled.
- But what does it mean? All orderings are equiprobable

Two practical questions

- How do we use probabilistic hypothesis in practice?
- How do we select a plausible probabilistic hypothesis?

Cournot principle

Cournot principle

- Philosophical: Events with negligible probabilities are impossible.

Cournot principle

- Philosophical: Events with negligible probabilities are impossible.
- Practical: Having two equally undesirable events, consider first the event that has greater probability

Cournot principle

- Philosophical: Events with negligible probabilities are impossible.
- Practical: Having two equally undesirable events, consider first the event that has greater probability

E. Borel: …Fewer than a million people live in Paris. Newspapers daily inform us about strange events or accidents that happen to some of them. Our life would be impossible if we were afraid of all adventures we read about. So one can say that from a practical viewpoint one can ignore with probability less than one over million… Often trying to avoid something bad we are confronted with even worse…

Cournot principle: testing hypotheses

Cournot principle: testing hypotheses

- A hypothesis should be rejected if it assigns a negligible probability to an event that has happened

Cournot principle: testing hypotheses

- A hypothesis should be rejected if it assigns a negligible probability to an event that has happened
- Needs some restriction – not all events can be used

Cournot principle: testing hypotheses

- A hypothesis should be rejected if it assigns a negligible probability to an event that has happened
- Needs some restriction – not all events can be used
- Practitioner: if a very unprobable event specified in advance has happened

Cournot principle: testing hypotheses

- A hypothesis should be rejected if it assigns a negligible probability to an event that has happened
- Needs some restriction – not all events can be used
- Practitioner: if a very unprobable event specified in advance has happened
- Mathematician: if a very unprobable simple event has happened

Frequency interpretation

Cournot's principle implies that frequency is close to probability

Frequency interpretation

Cournot's principle implies that frequency is close to probability

Bernoulli distribution B_p on n-bit sequences

Frequency interpretation

Cournot's principle implies that frequency is close to probability
Bernoulli distribution B_p on n-bit sequences
Event: $|\text{frequency} - p| > \varepsilon$ has small probability (and is simple)

Kolmogorov complexity

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$
- $l(p)$ — length of program p

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$
- $l(p)$ — length of program p
- U — an universal interpreter (that makes K minimal)

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$
- $l(p)$ — length of program p
- U — an universal interpreter (that makes K minimal)
- defined up to $O(1)$ additive term

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$
- $l(p)$ — length of program p
- U — an universal interpreter (that makes K minimal)
- defined up to $O(1)$ additive term
- $K(x) \leq l(x) + O(1)$

Kolmogorov complexity

- $K(x) = \min\{l(p) \mid U(p) = x\}$
- $l(p)$ — length of program p
- U — an universal interpreter (that makes K minimal)
- defined up to $O(1)$ additive term
- $K(x) \leq l(x) + O(1)$
- for most strings of length n the complexity is close to n: it is less than $n - d$ for 2^{-d}-fraction only

Is Kolmogorov complexity practical?

Is Kolmogorov complexity practical?

- defined only up to a constant, so the question "What is K(010001)" or "Which of the strings 0001 and 100 is simpler" has no sense

Is Kolmogorov complexity practical?

- defined only up to a constant, so the question "What is K(010001)" or "Which of the strings 0001 and 100 is simpler" has no sense
- reasonable interpreters give values that differ by several thousands, so the Kolmogorov complexity of human DNA is defined with $< 1\%$ error

Is Kolmogorov complexity practical?

- defined only up to a constant, so the question "What is K(010001)" or "Which of the strings 0001 and 100 is simpler" has no sense
- reasonable interpreters give values that differ by several thousands, so the Kolmogorov complexity of human DNA is defined with $< 1\%$ error
- Kolmogorov complexity is noncomputable; moreover, it has no computable lower bounds. So K(DNA) never will be known"

Is Kolmogorov complexity practical?

- defined only up to a constant, so the question "What is K(010001)" or "Which of the strings 0001 and 100 is simpler" has no sense
- reasonable interpreters give values that differ by several thousands, so the Kolmogorov complexity of human DNA is defined with $< 1\%$ error
- Kolmogorov complexity is noncomputable; moreover, it has no computable lower bounds. So $K(\mathrm{DNA})$ never will be known"
- Kolmogorov complexity does not take into account resources used by the program that generates x

Testing a hypothesis and Kolmogorov complexity

Testing a hypothesis and Kolmogorov complexity

fair coin (theory) string x (data)

Testing a hypothesis and Kolmogorov complexity

fair coin (theory) string x (data)

Do they match?

Testing a hypothesis and Kolmogorov complexity

fair coin (theory) string x (data)

Do they match?

Not yes/no-question; measure of disbelief, "randomness deficiency" $d(x)$

Testing a hypothesis and Kolmogorov complexity

fair coin (theory) string x (data)

Do they match?

Not yes/no-question; measure of disbelief,
"randomness deficiency" $d(x)$ $d(x) = l(x) - K(x)$

Testing a hypothesis and Kolmogorov complexity

fair coin (theory) string x (data)

Do they match?

Not yes/no-question; measure of disbelief, "randomness deficiency" $d(x)$ $d(x) = l(x) - K(x)$

randomness = incompressibility: we reject the hypothesis of fair coin if the observed string is compressible

Testing a hypothesis and Kolmogorov complexity

fair coin (theory)   string x (data)

Do they match?

Not yes/no-question; measure of disbelief, "randomness deficiency" $d(x)$ $d(x) = l(x) - K(x)$

randomness = incompressibility: we reject the hypothesis of fair coin if the observed string is compressible

for non-uniform distribution:
$d(x) = \log_2 P(x) - K(x)$

An incompressibility paradox

we do not think that fair coin never produces (or less frequently produces) compressible sequences, but they discredit the fairness hypothesis (unlike others)

⟨...⟩ the very Calculus of Probabilities to which I have referred, forbids all idea of the extension of the parallel ⟨...⟩ This is one of those anomalous propositions which, seemingly appealing to thought altogether apart from the mathematical, is yet one which only the mathematician can fully entertain. Nothing, for example, is more difficult than to convince the merely general reader that the fact of sixes having been thrown twice in succession by a player at dice, is sufficient cause for betting the largest odds that sixes will not be thrown in the third attempt. A suggestion to this effect is usually rejected by the intellect at once. It does not appear that the two throws which have been completed, and which lie now absolutely in the Past, can have influence upon the throw which exists only in the Future. The chance for throwing sixes seems to be precisely as it was at any ordinary time—that is to say, subject only to the influence of the various other throws which may be made by the dice. And this is a reflection which appears so exceedingly obvious that attempts to controvert it are received more frequently with a derisive smile than with any thing like respectful attention. The error here involved — a gross error redolent of mischief — I cannot pretend to expose within the limits assigned me at present. (Edgar Poe)

Infinite sequences

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)
- infinite string ω (data) of zeros and ones

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)
- infinite string ω (data) of zeros and ones

do they match?

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)
- infinite string ω (data) of zeros and ones

do they match?

yes/no-question; ω can be *Martin-Löf* random with respect to P or not

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)
- infinite string ω (data) of zeros and ones

do they match?

yes/no-question; ω can be *Martin-Löf* random with respect to P or not

it is random if randomness deficiencies of its prefixes are bounded

Infinite sequences

- Probability distribution P on the space of all infinite 0-1-sequences (theory)
- infinite string ω (data) of zeros and ones

do they match?

yes/no-question; ω can be *Martin-Löf* random with respect to P or not

it is random if randomness deficiencies of its prefixes are bounded [if randomness deficiency is defined in a proper way]

Returning to natural sciences

Returning to natural sciences

One can be practically sure that fair coin will never produce a sequence of 10^6 zeros and ones that can be zip-compressed at least by 1%.

Returning to natural sciences

One can be practically sure that fair coin will never produce a sequence of 10^6 zeros and ones that can be zip-compressed at least by 1%.

(Probability less than 2^{-1000})

Returning to natural sciences

One can be practically sure that fair coin will never produce a sequence of 10^6 zeros and ones that can be zip-compressed at least by 1%.

(Probability less than 2^{-1000})

Is this law of nature a consequence of mechanical laws?

Returning to natural sciences

One can be practically sure that fair coin will never produce a sequence of 10^6 zeros and ones that can be zip-compressed at least by 1%.

(Probability less than 2^{-1000})

Is this law of nature a consequence of mechanical laws?

Less philosophical version:

Returning to natural sciences

One can be practically sure that fair coin will never produce a sequence of 10^6 zeros and ones that can be zip-compressed at least by 1%.

(Probability less than 2^{-1000})

Is this law of nature a consequence of mechanical laws?

Less philosophical version:

Imagine we have a dice (nonsymmetric), know the position of its center of gravity, and have unlimited computation power. Can we compute probabilities of different outcomes using mechanical laws?

Probability in natural sciences

Probability in natural sciences

Phase space of a dice and a flow in this space

Probability in natural sciences

Phase space of a dice and a flow in this space

Mixing property: the neighborhood of the initial condition is mapped to outcomes $1 \ldots 6$, and preimages are densely mixed, so the conditional probabilities in a small neighborhood are the same for different neighborhoods and distributions

Probability in natural sciences

Phase space of a dice and a flow in this space

Mixing property: the neighborhood of the initial condition is mapped to outcomes $1 \ldots 6$, and preimages are densely mixed, so the conditional probabilities in a small neighborhood are the same for different neighborhoods and distributions

Theoretically they can be computed

Model example

Model example

Phase space: $[0, 1]$

Model example

Phase space: $[0, 1]$

Transformation: at each step x is transformed into $2x \mod 1$

Model example

Phase space: $[0, 1]$

Transformation: at each step x is transformed into $2x \mod 1$

We observe whether the current position is in the left half (0) or right half (1)

Model example

Phase space: $[0, 1]$

Transformation: at each step x is transformed into $2x \mod 1$

We observe whether the current position is in the left half (0) or right half (1)

In this way for initial condition x we get a sequence of observations: in which half is $x, T(x), T(T(x)), \ldots$

Model example

Phase space: $[0,1]$

Transformation: at each step x is transformed into $2x \mod 1$

We observe whether the current position is in the left half (0) or right half (1)

In this way for initial condition x we get a sequence of observations: in which half is $x, T(x), T(T(x)), \ldots$

the same kind of mixing property

Model example

Phase space: $[0, 1]$

Transformation: at each step x is transformed into $2x \bmod 1$

We observe whether the current position is in the left half (0) or right half (1)

In this way for initial condition x we get a sequence of observations: in which half is $x, T(x), T(T(x)), \ldots$

the same kind of mixing property

what happens: initial condition is revealed bit by bit

A new law of nature?

A new law of nature?

mixing does not create randomness but just reveals the randomness in the initial condition

A new law of nature?

mixing does not create randomness but just reveals the randomness in the initial condition

dynamical laws + one more: the world was created in an incompressible state

A new law of nature?

mixing does not create randomness but just reveals the randomness in the initial condition

dynamical laws + one more: the world was created in an incompressible state

this law together with mixing property implies that outcomes for a fair coin form an incompressible sequence

Pseudorandom number generators (Yao–Micali)

$$G : \mathbb{B}^{1000} \longrightarrow \mathbb{B}^{1000000}$$

Pseudorandom number generators (Yao–Micali)

$G : \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$

easily computable (polynomial time)

Pseudorandom number generators (Yao–Micali)

$G : \mathbb{B}^{1000} \longrightarrow \mathbb{B}^{1000000}$

easily computable (polynomial time)

random seed $\in \mathbb{B}^n$

Pseudorandom number generators (Yao–Micali)

$G : \mathbb{B}^{1000} \longrightarrow \mathbb{B}^{1000000}$

easily computable (polynomial time)

random seed $\in \mathbb{B}^n$

converted to "pseudorandom" G(seed)

Pseudorandom number generators (Yao–Micali)

$G : \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$

easily computable (polynomial time)

random seed $\in \mathbb{B}^n$

converted to "pseudorandom" $G(\text{seed})$

for every feasible test $T : \mathbb{B}^{1000000} \to \mathbb{B}$ the fraction of $s \in \mathbb{B}^{1000}$ such that $T(G(s)) =$ **True** almost coincides with the fraction of $r \in \mathbb{B}^{1000000}$ such that $T(r) =$ **True**

Pseudorandom number generators (Yao–Micali)

$G : \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$

easily computable (polynomial time)

random seed $\in \mathbb{B}^n$

converted to "pseudorandom" $G(\text{seed})$

for every feasible test $T : \mathbb{B}^{1000000} \to \mathbb{B}$ the fraction of $s \in \mathbb{B}^{1000}$ such that $T(G(s)) =$ **True** almost coincides with the fraction of $r \in \mathbb{B}^{1000000}$ such that $T(r) =$ **True**

"things seem random because we do not know they are not"

Thermodynamics

Thermodynamics

Second Law of Thermodynamics

Thermodynamics

Second Law of Thermodynamics

- entropy can only increase;

Thermodynamics

Second Law of Thermodynamics

- entropy can only increase;
- A perpetuum mobile of the second kind does not exist.

Thermodynamics

Second Law of Thermodynamics

- entropy can only increase;
- A perpetuum mobile of the second kind does not exist.

Usual remarks:

Thermodynamics

Second Law of Thermodynamics

- entropy can only increase;
- A perpetuum mobile of the second kind does not exist.

Usual remarks:

- these formulations are equivalent;

Thermodynamics

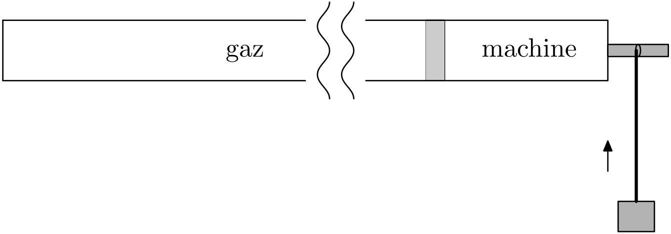
Second Law of Thermodynamics

- entropy can only increase;
- A perpetuum mobile of the second kind does not exist.

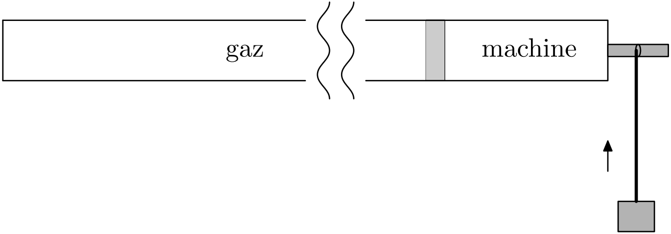
Usual remarks:

- these formulations are equivalent;
- the first one cannot be a corollary of dynamic laws since it is not time-symmetric

Perpetuum mobile of the second kind



Perpetuum mobile of the second kind



moves the weight arbitrary high if the reservoir is large enough (for most states of the gaz in the reservoir)

“Proof” of impossibility

“Proof” of impossibility

Phase space is almost a product $S_1 \times S_2$

“Proof” of impossibility

Phase space is almost a product $S_1 \times S_2$

Invariant measure on the phase space:

"Proof" of impossibility

Phase space is almost a product $S_1 \times S_2$

Invariant measure on the phase space:

initial condition: more energy in gaz; final condition: more energy in the weight

“Proof” of impossibility

Phase space is almost a product $S_1 \times S_2$

Invariant measure on the phase space:

initial condition: more energy in gaz; final condition: more energy in the weight

Volume in S_1 depends on T much more than in S_2 (# of degrees of freedom)

"Proof" of impossibility

Phase space is almost a product $S_1 \times S_2$

Invariant measure on the phase space:

initial condition: more energy in gaz; final condition: more energy in the weight

Volume in S_1 depends on T much more than in S_2 (# of degrees of freedom)

Large set cannot be mapped into a small one

Quantum mechanics

Quantum mechanics

common wisdom: “unlike statistical mechanics, which is microscopically deterministic, the quantum mechanics has intrinsic nondeterminism (randomness)”

Quantum mechanics

common wisdom: "unlike statistical mechanics, which is microscopically deterministic, the quantum mechanics has intrinsic nondeterminism (randomness)"

random coin vs. radioactive decay

Quantum mechanics

common wisdom: "unlike statistical mechanics, which is microscopically deterministic, the quantum mechanics has intrinsic nondeterminism (randomness)"

random coin vs. radioactive decay

q-Cournot principle: the events with negligible amplitude do not happen