Individual random objects in computer science and ‘real life’

Alexander Shen, LIRMM,
CNRS & University of Montpellier,
on leave from IITP RAS, Moscow

Hilbert Sixth Problem conference,
Leicester, May 3, 2016

Disclaimer

- More (old) questions than (new) answers
- More philosophy than theorems
- Just a series of examples to think about

Disclaimer

- More (old) questions than (new) answers
- More philosophy than theorems
- Just a series of examples to think about

Disclaimer

- More (old) questions than (new) answers
- More philosophy than theorems
- Just a series of examples to think about

Disclaimer

- More (old) questions than (new) answers
- More philosophy than theorems
- Just a series of examples to think about

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Introduction: law of incompressibility

- Experiment: toss a coin 80000 times, then apply zip compressor to 10000 bytes obtained
- Claim: more than 1% compression is not possible
- 1% compression: $10000 \times 8 \to 9900 \times 8$
- there are at most $2 \times 2^{9900 \times 8}$ files of length ≤ 9900
- at most $2 \times 2^{9900 \times 8}$ 1%-compressible files of length 10000
- about 2^{-799}-fraction = impossibility
- as reliable as Ohm's law (or any other)
- does it follow from known physics' laws? if yes, how?

Classical mechanics: why a fair coin is fair?

- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \ldots p_6$
- in principle is solvable numerically

Classical mechanics: why a fair coin is fair?

- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \ldots p_6$
- in principle is solvable numerically

Classical mechanics: why a fair coin is fair?

- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \dots p_6$
- in principle is solvable numerically

Classical mechanics: why a fair coin is fair?

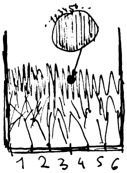
- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \ldots p_6$
- in principle is solvable numerically

Classical mechanics: why a fair coin is fair?

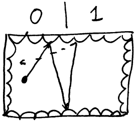
- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \ldots p_6$
- in principle is solvable numerically

Classical mechanics: why a fair coin is fair?

- can we *prove* that a fair coin is indeed fair using mechanics' laws?
- physics is more about computations than proofs
- better question: a dice shape and center of gravity are known; compute $p_1 \dots p_6$
- in principle is solvable numerically
- phase space is split into six rather dense sets; relative measure of each inside a not very small volume should be almost constant

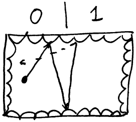


Classical mechanics: a point in a billiard



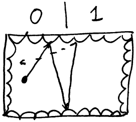
- consider a particle in a billiard with some initial condition
- and register its position after time $T, 2T, 3T, \ldots$ for some large constant T
- $0/1 =$ left half / right half
- get a bit sequence that we expect to be 'random'

Classical mechanics: a point in a billiard



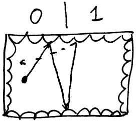
- consider a particle in a billiard with some initial condition
- and register its position after time $T, 2T, 3T, \ldots$ for some large constant T
- $0/1$ = left half / right half
- get a bit sequence that we expect to be 'random'

Classical mechanics: a point in a billiard



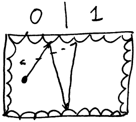
- consider a particle in a billiard with some initial condition
- and register its position after time $T, 2T, 3T, \ldots$ for some large constant T
- 0/1 = left half / right half
- get a bit sequence that we expect to be 'random'

Classical mechanics: a point in a billiard



- consider a particle in a billiard with some initial condition
- and register its position after time $T, 2T, 3T, \ldots$ for some large constant T
- $0/1$ = left half / right half
- get a bit sequence that we expect to be 'random'

Classical mechanics: a point in a billiard



- consider a particle in a billiard with some initial condition
- and register its position after time $T, 2T, 3T, \ldots$ for some large constant T
- $0/1$ = left half / right half
- get a bit sequence that we expect to be 'random'

Classical mechanics: revealing randomness

- model system: $T : x \in [0, 1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \dots$ (left or right half)
- initial condition: real $x = x_0x_1x_2\dots$ produces bits $x_0, x_1, x_2, \dots$
- $\dots$ just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T : x \in [0, 1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0 x_1 x_2 \ldots$ produces bits $x_0, x_1, x_2, \ldots$
- $\ldots$ just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T\colon x \in [0,1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0x_1x_2\ldots$ produces bits $x_0, x_1, x_2, \ldots$
- $\ldots$ just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T : x \in [0, 1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0 x_1 x_2 \ldots$ produces bits x_0, x_1, x_2, . . .
- . . . just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T\colon x \in [0,1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0x_1x_2\ldots$ produces bits $x_0,\ x_1,\ x_2, \ldots$
- …just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T : x \in [0,1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0 x_1 x_2 \ldots$ produces bits $x_0,\ x_1,\ x_2, \ldots$
- ...just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Classical mechanics: revealing randomness

- model system: $T : x \in [0,1] \mapsto 2x \bmod 1$
- the position of $x, T(x), T(T(x)), T(T(T(x))), \ldots$ (left or right half)
- initial condition: real $x = x_0 x_1 x_2 \ldots$ produces bits $x_0, x_1, x_2, \ldots$
- ...just reveals bits of x
- initial condition as a source of randomness
- some dynamic systems reveal the randomness hidden in the initial condition (while other do not)

Real life: tables of random numbers

- you buy a book with table of random numbers

2 TABLE OF RANDOM DIGITS

00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813

- you see page filled with zeros
- you complain: "look, this combination has astronomically small probability"
- but the same is true for any other combination of digits — answers the publisher
- how do you justify your complaint?

Real life: tables of random numbers

- you buy a book with table of random numbers

2 TABLE OF RANDOM DIGITS

00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813

- you see page filled with zeros
- you complain: “look, this combination has astronomically small probability”
- but the same is true for any other combination of digits — answers the publisher
- how do you justify your complaint?

Real life: tables of random numbers

- you buy a book with table of random numbers

2 TABLE OF RANDOM DIGITS

00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813

- you see page filled with zeros
- you complain: "look, this combination has astronomically small probability"
- but the same is true for any other combination of digits — answers the publisher
- how do you justify your complaint?

Real life: tables of random numbers

- you buy a book with table of random numbers

2 TABLE OF RANDOM DIGITS

00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813

- you see page filled with zeros
- you complain: "look, this combination has astronomically small probability"
- but the same is true for any other combination of digits — answers the publisher
- how do you justify your complaint?

Real life: tables of random numbers

- you buy a book with table of random numbers

2 TABLE OF RANDOM DIGITS

00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813

- you see page filled with zeros
- you complain: “look, this combination has astronomically small probability”
- but the same is true for any other combination of digits — answers the publisher
- how do you justify your complaint?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Real life: randomness of individual objects?

- multiple choice test (twenty A/B questions)
- order of answers randomized before printing each copy (A/B are exchanged randomly)
- in some copy all correct answers happen to be A
- should it be used?
- one more example: a factory that produces preshuffled deck of cards
- quality control takes one deck to check it is OK
- what should it check?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with ‘real world’?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: ‘fair coin’ hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with 'real world'?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: 'fair coin' hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability theory

- probability theory as part of measure theory that deals with independence: no problems
- what is the relation with ‘real world’?
- observation — statistical model (probability distribution) — recommendations —...
- example of a model: ‘fair coin’ hypothesis ("head and tail have probability 1/2")
- what does it mean?
- what this hypothesis predicts?
- how it can be rejected experimentally?

Foundations of probability: frequency approach

- probability 1/2: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability 1/2: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability 1/2: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability $1/2$: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability 1/2: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability $1/2$: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability $1/2$: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: frequency approach

- probability $1/2$: what does it mean?
- if we toss the coin many times, tails and heads appear equally often
- exactly?
- no, but large deviations happen rarely: difference more than $10\sqrt{N}$ for N coin tossings is unlikely
- unlikely?
- yes, this happens with small probability
- probability??

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: …je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: ...je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: …je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: …je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: ... je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle I

- how to break this circle?
- Cournot principle: events with very small probability do not happen
- Borel: …je suis arrivé à la conclusion qu'on ne devrait pas craindre d'employer le mot de *certitude* pour désigner une probabilité qui differe de l'unité d'une quantité suffisamment petite
- more precisely, "other things equal, you should worry more about more probable events"
- Borel: "Souvent la peur d'un mal fait tomber dans un pire. Pour savoir distinguer le pire, il est bon de connaître les probabilités des diverses éventualités"

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

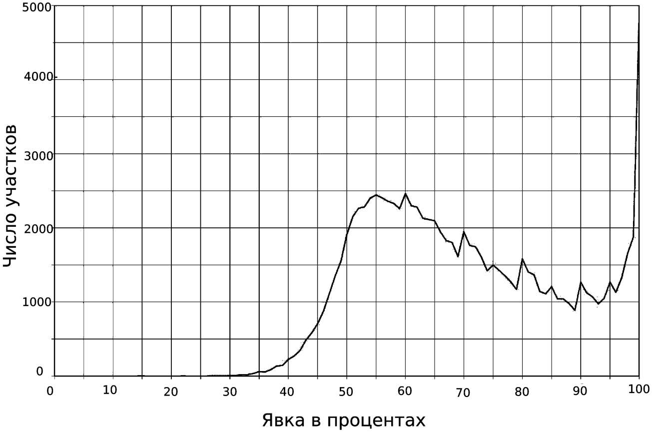
Foundations of probability: Cournot principle II

- recall the question about random digits table
- seeing zeros we say that an event that has negligible probability (under the hypothesis) happened; so the hypothesis is rejected
- but what about the other combinations?
- why we do not reject the hypothesis seeing some other combination?
- "if a simple event with negligible probability under the hypothesis happens, reject the hypothesis"
- simple or specified in advance?
- what is simple?

Borel on hypotheses' testing

Consider a random integer between 1 000 000 and 2 000 000. The probability that it is equal to 1342517, is one over million; the probability that it is equal to 1500000, is also one over million.
…When a number like this appears as an angle measured in centesimal seconds, we do not ask ourselves what is the probability that this angle is exactly $13^\circ 42' 51'',7$ because we never would be interested in such a question before the measurement. Of course, the angle should have some value, and whatever this value is (up to a tenth of a second), we may measure it and say that the *a priori* probability to get this value is one in ten millions, so an extraordinary event has happened…
The quest is whether the same reservations apply if one of the angles formed by three starts has a *remarkable* value, for example, is equal to the angle in the equilateral triange… or the half of the right angle… What can we say about that? one should try hard to avoid the temptation to consider some event not fixed *before the experiment*, as a *remarkable* one, because a lot of events could look remarkable from some viewpoint.

Digression: real life (I, 2007)



[число участков = number of polling stations
явка в процентах = percentage of voters that participated in the vote]

Digression: real life (II, 2014)

Registered voters: 306258
Participated in the vote: 274101
Voted for: 262041

$274101/306258 = 0.895000294$
$262041/274101 = 0.95600161$

$0.895 * 306258 = 274100.91$
$0.956 * 274101 = 262040.556$

Other examples:
http://kireev.livejournal.com/1095568.html

Digression: real life (II, 2014)

Registered voters: 306258
Participated in the vote: 274101
Voted for: 262041

$274101/306258 = 0.895000294$
$262041/274101 = 0.95600161$

$0.895 * 306258 = 274100.91$
$0.956 * 274101 = 262040.556$

Other examples:
http://kireev.livejournal.com/1095568.html

Digression: real life (II, 2014)

Registered voters: 306258
Participated in the vote: 274101
Voted for: 262041

$274101/306258 = 0.895000294$
$262041/274101 = 0.95600161$

$0.895 * 306258 = 274100.91$
$0.956 * 274101 = 262040.556$

Other examples:
http://kireev.livejournal.com/1095568.html

Digression: real life (II, 2014)

Registered voters: 306258
Participated in the vote: 274101
Voted for: 262041

$274101/306258 = 0.895000294$
$262041/274101 = 0.95600161$

$0.895 * 306258 = 274100.91$
$0.956 * 274101 = 262040.556$

Other examples:
http://kireev.livejournal.com/1095568.html

Digression: real life (II, 2014)

Registered voters: 306258
Participated in the vote: 274101
Voted for: 262041

$274101/306258 = 0.895000294$
$262041/274101 = 0.95600161$

$0.895 * 306258 = 274100.91$
$0.956 * 274101 = 262040.556$

Other examples:
http://kireev.livejournal.com/1095568.html

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- "individual random bit strings"
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- “individual random bit strings”
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- "individual random bit strings"
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- "individual random bit strings"
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- "individual random bit strings"
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: individual random objects

- which bit strings do not convince us to reject the hypothesis of a fair coin?
- "individual random bit strings"
- suggested answer: incompressible
- there is no short program (much shorten than the string itself) that produced the string
- formally: a string x is random if its Kolmogorov complexity $C(x)$ is close to its length

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
  - proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
  - proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \le C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Algorithmic information theory: Kolmogorov complexity

- Consider some programming language (binary strings as programs and outputs; no input)
- Let D be an interpreter of this language
- $C_D(x) = \min\{l(p) \mid D(p) = x\}$
- (= minimal length of a program that outputs x)
- depends on D
- D is better than D' if $C_D(x) \leq C_{D'}(x) + c$ for some c and all x
- there exists optimal D that is better that any other algorithm D'
- proof: let D be a universal programming language that can simulate any other one

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations

- fix some optimal D and call $C_D(x)$ "Kolmogorov complexity" of x
- defined up to $O(1)$ additive term
- "what is more complex: 000111000111 or 010101010101": a meaningless question
- $C(x) \leq l(x) + O(1)$
- for most strings of length n we have $C(x) \approx n$
- more precisely, $C(x) < n - d$ for at most 2^{-d}-fraction of n-bit strings
- most strings are incompressible (random)

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Kolmogorov complexity: properties and limitations II

- C has natural properties the measure of information should have
- for example $C(A(x)) \leq C(x) + O(1)$ for algorithmic transformation A
- $O(1)$-constant depends on A
- bad news: C not computable
- even no computable lower bounds
- Gödel–Chaitin: statement of the form $C(x) > n$ (for specific x and n) are never provable for large enough values of n (though most are true)
- Does not take into account the resources used to produce x

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy

- n possible messages in a channel
- probabilities (frequencies) $p_1, \ldots, p_n$
- want to develop a uniquely decodable code for these messages
- to minimize the average length, frequent messages should have shorter code
- Shannon: lower bound $H(p_1, \ldots, p_n)$ for uniquely decodable codes
- can be (almost) achieved by prefix codes
- $H(p_1, \ldots, p_n) = \sum p_i \log(1/p_i)$

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of “Hamlet”: meaningless
- Kolmogorov complexity of “Hamlet”: meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Shannon entropy and Kolmogorov complexity

- Shannon entropy requires a probability distribution
- entropy per letter in an English text: not so well defined
- entropy of "Hamlet": meaningless
- Kolmogorov complexity of "Hamlet": meaningful, no hope to answer
- closely related: for a random source the Kolmogorov complexity of the output is close to Shannon entropy
- closely related: the same linear inequalities

Example: inequalities and equalities

- $H(\xi,\eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x,y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi,\eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x,y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x,y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi,\eta) \le H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x,y) \le C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi,\eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x,y) = C(x) + C(y|x) + \ldots$
- $\le$: concatenation
- $\ge$: non-trivial: why the shortest program for (x,y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi,\eta) \le H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x,y) \le C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi,\eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x,y) = C(x) + C(y|x) + \dots$
- $\le$: concatenation
- $\ge$: non-trivial: why the shortest program for (x,y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \dots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi, \eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x, y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi, \eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x, y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x, y) should start by specifying x?

Example: inequalities and equalities

- $H(\xi,\eta) \leq H(\xi) + H(\eta)$
- proof: convexity of logarithm
- $C(x,y) \leq C(x) + C(y) + O(\log(|x| + |y|))$
- proof: concatenate the programs using some separator
- $H(\xi,\eta) = H(\xi) + H(\eta|\xi)$
- proof: by definition
- $C(x,y) = C(x) + C(y|x) + \ldots$
- $\leq$: concatenation
- $\geq$: non-trivial: why the shortest program for (x,y) should start by specifying x?

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G : \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- 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": pseudoentropy.
- PRNG exists if NP$\neq$P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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": pseudoentropy.
- PRNG exists if NP≠P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G : \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- 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": pseudoentropy.
- PRNG exists if NP≠P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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": pseudoentropy.
- PRNG exists if NP≠P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ the fraction of $s \in \mathbb{B}^{1000}$ such that $T(G(s)) = \mathbf{True}$ almost coincides with the fraction of $r \in \mathbb{B}^{1000000}$ such that $T(r) = \mathbf{True}$
- "things seem random because we do not know they are not": pseudoentropy.
- PRNG exists if NP≠P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but “indistinguishable from random”...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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”: pseudoentropy.
- PRNG exists if NP≠P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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": pseudoentropy.
- PRNG exists if NP$\neq$P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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": pseudoentropy.
- PRNG exists if NP$\neq$P and, moreover, one-way functions exist

Randomness as our ignorance?

- pseudorandom number generators (Yao–Micali)
- $G\colon \mathbb{B}^{1000} \to \mathbb{B}^{1000000}$
- easily computable (polynomial time)
- random 1000-bit seed converted to 10^6-bit pseudorandom string: not random (compressible) but "indistinguishable from random"...
- for every feasible test $T\colon \mathbb{B}^{1000000} \to \mathbb{B}$ 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": pseudoentropy.
- PRNG exists if NP$\neq$P and, moreover, one-way functions exist

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

Entropy in physics

- Imagine some system, e.g., ideal gas
- ‘Second Law: ‘entropy increases”
- does it mean that entropy is a function of state (on phase space)?
- how is it compatible with time symmetry?
- can the Second Law be derived from other laws?
- is it equivalent to the nonexistent of a perpetuum mobile of the second kind?
- is entropy in physics more like Shannon entropy or Kolmogorov complexity?

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

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

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

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

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

Thermodynamics (a layman's view)

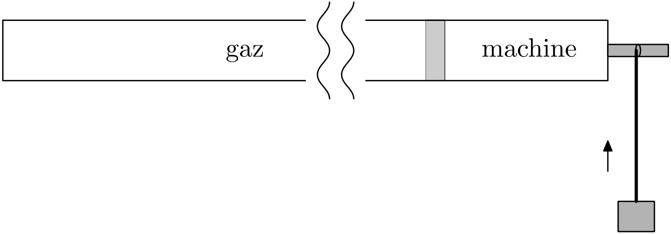
Second Law of Thermodynamics

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

Usual remarks:

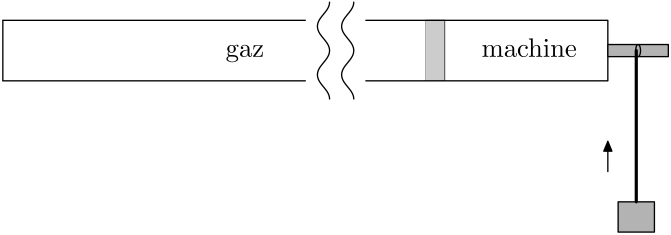
- these formulations are equivalent;
- the first one cannot be a corollary of dynamic laws since it is not time-symmetric

Perpetuum mobile of the second kind



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

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

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

"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

“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

“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

"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

“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