Online probability, complexity and randomness

Alexey Chernov (Royal Holloway College),
Alexander Shen (LIF, Marseille)[1],
Nikolay Vereshchagin (Moscow University),
Vladimir Vovk (Royal Holloway College)

ALT 2008 (October 2008)

[1] on leave from IITP, Moscow

Algorithmic randomness:

$$\boxed{\text{Fair Coin}} \longrightarrow 0\ 1\ 1\ 0\ 0\ 1 \ldots$$

Algorithmic randomness:

$$\boxed{\text{Fair Coin}} \rightarrow 0\ 1\ 1\ 0\ 0\ 1 \ldots$$

Some sequences look suspicious:

$$0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \ldots$$

Algorithmic randomness:

$$\boxed{\text{Fair Coin}} \to 0\ 1\ 1\ 0\ 0\ 1 \ldots$$

Some sequences look suspicious:

$$0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \ldots$$

or

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

Algorithmic randomness:

$$\boxed{\text{Fair Coin}} \to 0\ 1\ 1\ 0\ 0\ 1 \ldots$$

Some sequences look suspicious:

$$0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \ldots$$

or

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

but not all

$\boxed{\text{Statistical Hypothesis}} \rightarrow$ observed behavior

$\boxed{\text{Statistical Hypothesis}} \longrightarrow$ observed behavior

compatible or not?

$\boxed{\text{Statistical Hypothesis}} \rightarrow$ observed behavior

compatible or not?

Probability distribution P sequence α

$\boxed{\text{Statistical Hypothesis}} \rightarrow$ observed behavior

compatible or not?

Probability distribution P sequence α

Martin-Löf: sequence α may be *random* or *non-random* with respect to P

$\boxed{\begin{array}{c}\text{Statistical}\\ \text{Hypothesis}\end{array}} \rightarrow$ observed behavior

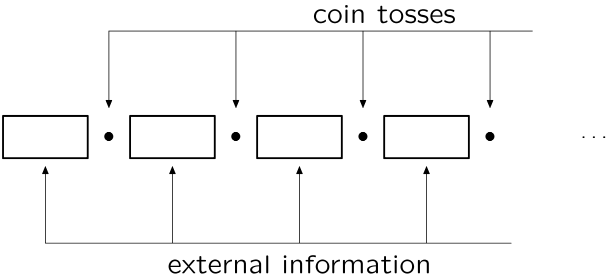
compatible or not?

Probability distribution P sequence α

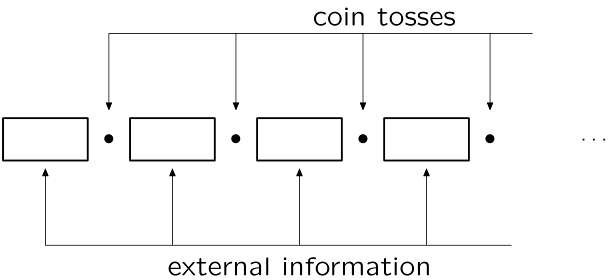
Martin-Löf: sequence α may be *random* or *non-random* with respect to P

(P is a computable distribution on the Cantor space $\{0,1\}^\infty$ of binary sequences)

"Real" life:

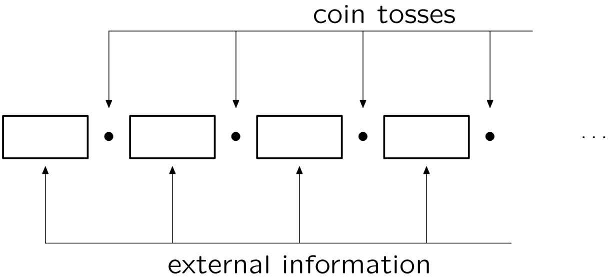


“Real” life:



Asking whether bits look random, we should take context into account

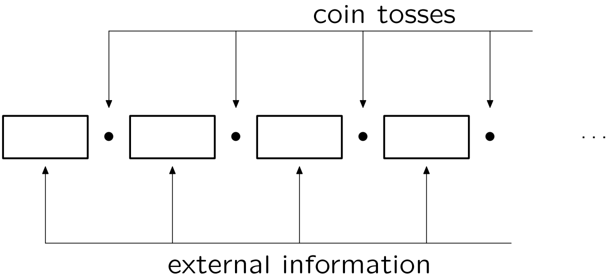
"Real" life:



Asking whether bits look random, we should take context into account

lottery result = f(yesterday newspaper)?

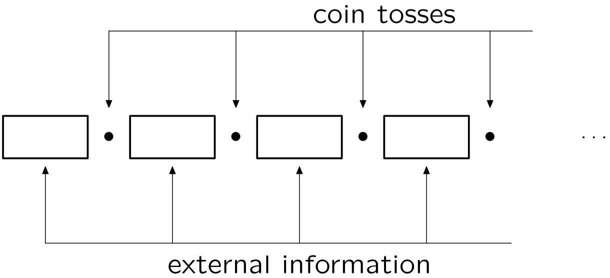
"Real" life:



Asking whether bits look random, we should take context into account

lottery result = f(yesterday newspaper)?   BAD

“Real” life:

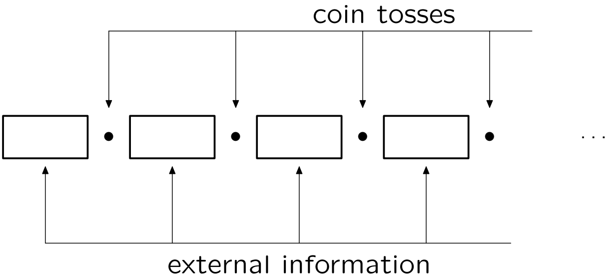


Asking whether bits look random, we should take context into account

lottery result = f(yesterday newspaper)?   BAD

lottery result = f(tomorrow newspaper)?

"Real" life:



Asking whether bits look random, we should take context into account

lottery result = f(yesterday newspaper)? BAD

lottery result = f(tomorrow newspaper)? OK

"On-line randomness": consider a sequence

$$x_1, b_1, x_2, b_2, x_3, b_3, \ldots$$

where x_i are strings and b_i bits.

"On-line randomness": consider a sequence

$$x_1, b_1, x_2, b_2, x_3, b_3, \ldots$$

where x_i are strings and b_i bits.

We may ask whether bits b_i are random in this sequence

"On-line randomness": consider a sequence

$$x_1, b_1, x_2, b_2, x_3, b_3, \ldots$$

where x_i are strings and b_i bits.

We may ask whether bits b_i are random in this sequence

OLR inbetween classical notions:

"On-line randomness": consider a sequence

$$x_1, b_1, x_2, b_2, x_3, b_3, \ldots$$

where x_i are strings and b_i bits.

We may ask whether bits b_i are random in this sequence

OLR inbetween classical notions:

OLR $\Rightarrow b_1, b_2, b_3, \ldots$ is ML-random;

"On-line randomness": consider a sequence

$$x_1, b_1, x_2, b_2, x_3, b_3, \ldots$$

where x_i are strings and b_i bits.

We may ask whether bits b_i are random in this sequence

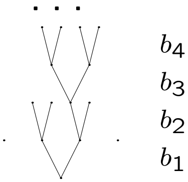
OLR inbetween classical notions:

OLR $\Rightarrow b_1, b_2, b_3, \ldots$ is ML-random;

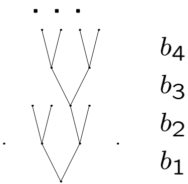
OLR $\Leftarrow b_1, b_2, b_3, \ldots$ is ML-random with oracle $x_1, x_2, x_3, \ldots$.

Martin-Löf randomness with respect to a distribution P (on binary sequences).

Martin-Löf randomness with respect to a distribution P (on binary sequences).



Martin-Löf randomness with respect to a distribution P (on binary sequences).

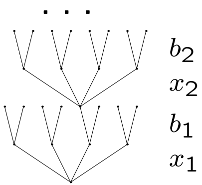


Such a distribution can be defined by conditional probabilities

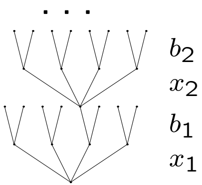
$$\Pr[b_i = 1 | b_1, b_2, \ldots, b_{i-1}]$$

On-line randomness is defined with respect to an on-line distribution P (on sequences $x_1, b_1, x_2, b_2, \ldots$).

On-line randomness is defined with respect to an on-line distribution P (on sequences $x_1, b_1, x_2, b_2, \ldots$).



On-line randomness is defined with respect to an on-line distribution P (on sequences $x_1, b_1, x_2, b_2, \ldots$).



Such a distribution can be defined by conditional probabilities

$$\Pr[b_i = 1 | x_1, b_1, x_2, b_2, \ldots, b_{i-1}, x_i]$$

On-line probability distribution $\longrightarrow$ a class of all distributions compatible with conditional probabilities

On-line probability distribution $\to$ a class of all distributions compatible with conditional probabilities

a notion of randomness with respect to a class of distributions (Levin, Gacs) [more details in the next talk]

On-line probability distribution $\rightarrow$ a class of all distributions compatible with conditional probabilities

a notion of randomness with respect to a class of distributions (Levin, Gacs) [more details in the next talk]

rather special class of distributions: other notions of algorithmic information theory can be generalized for the on-line framework

Decision complexity of a bit string $b_1, b_2, \ldots, b_n$: the minimal length of a program that prints (sequentially) $b_1, b_2, \ldots, b_n$ (and, may be, something else).

Decision complexity of a bit string $b_1, b_2, \ldots, b_n$: the minimal length of a program that prints (sequentially) $b_1, b_2, \ldots, b_n$ (and, may be, something else).

notation: $KR(b_1, b_2, \ldots, b_n)$

Decision complexity of a bit string $b_1, b_2, \ldots, b_n$: the minimal length of a program that prints (sequentially) $b_1, b_2, \ldots, b_n$ (and, may be, something else).

notation: $KR(b_1, b_2, \ldots, b_n)$

On-line decision complexity of $x_1, b_1, \ldots, x_n, b_n$: the minimal length of a program that reads x_1, then outputs b_1, then reads x_2, then outputs b_2, etc.

Decision complexity of a bit string $b_1, b_2, \ldots, b_n$: the minimal length of a program that prints (sequentially) $b_1, b_2, \ldots, b_n$ (and, may be, something else).

notation: $KR(b_1, b_2, \ldots, b_n)$

On-line decision complexity of $x_1, b_1, \ldots, x_n, b_n$: the minimal length of a program that reads x_1, then outputs b_1, then reads x_2, then outputs b_2, etc.

notation: $KR(x_1 \to b_1, x_2 \to b_2, \ldots, x_n \to b_n)$

A priori probability of a bit string $b_1, b_2, \ldots, b_n$: the probability that a universal probabilistic machine produces output bits $b_1, b_2, \ldots, b_n$ (and may be something else after that).

A priori probability of a bit string $b_1, b_2, \ldots, b_n$: the probability that a universal probabilistic machine produces output bits $b_1, b_2, \ldots, b_n$ (and may be something else after that).

(Universal machine emulates any other with positive probability. A priori probability is defined up to a multiplicative constant.)

A priori probability of a bit string $b_1, b_2, \ldots, b_n$: the probability that a universal probabilistic machine produces output bits $b_1, b_2, \ldots, b_n$ (and may be something else after that).

(Universal machine emulates any other with positive probability. A priori probability is defined up to a multiplicative constant.)

On-line *a priori* probability of $x_1, b_1, x_2, b_2, \ldots, x_n, b_n$ is a probability that a universal probabilistic machine, getting x_1 as input, produces b_1, then getting x_2 as input, produces b_2, etc.

Martingales:

- Casino produced bits $b_1, b_2, \ldots$ and announces the distribution: $\Pr[b_{n+1} = 1 | b_1, b_2, \ldots, b_n]$

Martingales:

- Casino produced bits $b_1, b_2, \ldots$ and announces the distribution: $\Pr[b_{n+1} = 1 | b_1, b_2, \ldots, b_n]$
- Player starts with some initial capital (no debt)

Martingales:

- Casino produced bits $b_1, b_2, \dots$ and announces the distribution: $\Pr[b_{n+1} = 1 | b_1, b_2, \dots, b_n]$
- Player starts with some initial capital (no debt)
- The game is fair: if declared probability of some outcome is p, then the bet on this outcome is multiplied by $1/p$

Martingales:

- Casino produced bits $b_1, b_2, \ldots$ and announces the distribution: $\Pr[b_{n+1} = 1 | b_1, b_2, \ldots, b_n]$
- Player starts with some initial capital (no debt)
- The game is fair: if declared probability of some outcome is p, then the bet on this outcome is multiplied by $1/p$
- Player's strategy can be described by a function $m(b_1, \ldots, b_n) =$ the capital after bits $b_1, \ldots, b_n$

Martingales:

- Casino produced bits $b_1, b_2, \ldots$ and announces the distribution: $\Pr[b_{n+1} = 1 | b_1, b_2, \ldots, b_n]$
- Player starts with some initial capital (no debt)
- The game is fair: if declared probability of some outcome is p, then the bet on this outcome is multiplied by $1/p$
- Player's strategy can be described by a function $m(b_1, \ldots, b_n)$ = the capital after bits $b_1, \ldots, b_n$
- This function is a martingale, i.e.,

$$m(b_1 \ldots b_n) = \Pr[b_{n+1} = 0 | b_1 \ldots b_n] m(b_1 \ldots b_n 0) + \\ + \Pr[b_{n+1} = 1 | b_1 \ldots b_n] m(b_1 \ldots b_n 1).$$

On-line martingales:

- Between the bits for betting some other activity happens in the Casino; the protocol is $x_1, b_1, x_2, b_2, \ldots$

On-line martingales:

- Between the bits for betting some other activity happens in the Casino; the protocol is $x_1, b_1, x_2, b_2, \ldots$
- Casino announces conditional probability only for b_i (on-line distribution): $\Pr[b_{n+1} = 1 | x_1, b_1, \ldots, x_n, b_n, x_{n+1}]$

On-line martingales:

- Between the bits for betting some other activity happens in the Casino; the protocol is $x_1, b_1, x_2, b_2, \ldots$
- Casino announces conditional probability only for b_i (on-line distribution): $\Pr[b_{n+1} = 1 | x_1, b_1, \ldots, x_n, b_n, x_{n+1}]$
- Player can make bets only on b_i

On-line martingales:

- Between the bits for betting some other activity happens in the Casino; the protocol is $x_1, b_1, x_2, b_2, \dots$
- Casino announces conditional probability only for b_i (on-line distribution): $\Pr[b_{n+1} = 1 | x_1, b_1, \dots, x_n, b_n, x_{n+1}]$
- Player can make bets only on b_i
- The game is fair

On-line martingales:

- Between the bits for betting some other activity happens in the Casino; the protocol is $x_1, b_1, x_2, b_2, \ldots$
- Casino announces conditional probability only for b_i (on-line distribution): $\Pr[b_{n+1} = 1 | x_1, b_1, \ldots, x_n, b_n, x_{n+1}]$
- Player can make bets only on b_i
- The game is fair
- Player's strategy can be described by a function $m(x_1, b_1, \ldots) =$ the capital after $x_1, b_1, \ldots$

- This function is a on-line martingale:

$$m(x_1, b_1, \ldots, x_n, b_n) = m(x_1, b_1, \ldots, x_n, b_n, x_{n+1})$$

(no bets on x_{n+1})

- This function is a on-line martingale:

$$m(x_1, b_1, \ldots, x_n, b_n) = m(x_1, b_1, \ldots, x_n, b_n, x_{n+1})$$

(no bets on x_{n+1})

- Betting is fair:

$$\begin{aligned} m(\ldots x_n, b_n, x_{n+1}) = \\ & \Pr[b_{n+1} = 0 | \ldots x_n, b_n, x_{n+1}] m(\ldots, x_n, b_n, x_{n+1}, 0) + \\ & + \Pr[b_{n+1} = 1 | \ldots x_n, b_n, x_{n+1}] m(\ldots, x_n, b_n, x_{n+1}, 1) \end{aligned}$$

Martingales and probability:

Martingales and probability:

Let P be a distribution on n-bit sequences $b_1, \ldots, b_n$

Martingales and probability:

Let P be a distribution on n-bit sequences $b_1, \ldots, b_n$

Let E be an event (a set of n-bit sequences)

Martingales and probability:

Let P be a distribution on n-bit sequences $b_1, \ldots, b_n$

Let E be an event (a set of n-bit sequences)

Ville's theorem: $\Pr[E]$ is the minimal initial capital needed for a martingale to achieve 1 on all elements of E

Martingales and probability:

Let P be a distribution on n-bit sequences $b_1, \ldots, b_n$

Let E be an event (a set of n-bit sequences)

Ville's theorem: $\Pr[E]$ is the minimal initial capital needed for a martingale to achieve 1 on all elements of E

In other terms, $1/\Pr[E]$ is the "market value" for the right to start playing with initial capital 1 and the insider information "outcome will be in E"

On-line

martingales and upper probability:

On-line

martingales and upper probability:

P: an on-line distribution on sequences $x_1 b_1 \ldots x_n b_n$;

On-line

martingales and upper probability:

P: an on-line distribution on sequences $x_1 b_1 \dots x_n b_n$;

E: an event (a set of sequences)

On-line

martingales and upper probability:

P: an on-line distribution on sequences $x_1 b_1 \ldots x_n b_n$;
E: an event (a set of sequences)

Consider the minimal initial capital needed for an on-line martingale to achieve 1 on all elements of E.
It can be called *upper probability* of E.

On-line

martingales and upper probability:

P: an on-line distribution on sequences $x_1 b_1 \dots x_n b_n$;
E: an event (a set of sequences)

Consider the minimal initial capital needed for an on-line martingale to achieve 1 on all elements of E. It can be called *upper probability* of E.

Upper probability is the maximal probability of E (maximum is taken over all distributions compatible with the on-line conditional probabilities)

On-line

martingales and upper probability:

P: an on-line distribution on sequences $x_1 b_1 \ldots x_n b_n$;
E: an event (a set of sequences)

Consider the minimal initial capital needed for an on-line martingale to achieve 1 on all elements of E. It can be called *upper probability* of E.

Upper probability is the maximal probability of E (maximum is taken over all distributions compatible with the on-line conditional probabilities)

Game: you choose x_i while b_i are generated with given (conditional) probabilities; you win if the outcome belongs to E. The winning probability is upper probability of E.

"Cournot principle": events with negligible probabilities never happen

"Cournot principle": events with negligible probabilities never happen

A short form of saying that:

- After a statistical hypothesis (a distribution) is accepted, one should be more aware of events that have bigger probability. (Corollary: events with negligible probabilities could be ignored.)

"Cournot principle": events with negligible probabilities never happen

A short form of saying that:

- After a statistical hypothesis (a distribution) is accepted, one should be more aware of events that have bigger probability. (Corollary: events with negligible probabilities could be ignored.)
- If a *simple* event that has negligible probability nevertheless happens, the statistical hypothesis should be rejected.

"Cournot principle": events with negligible probabilities never happen

A short form of saying that:

- After a statistical hypothesis (a distribution) is accepted, one should be more aware of events that have bigger probability. (Corollary: events with negligible probabilities could be ignored.)
- If a *simple* event that has negligible probability nevertheless happens, the statistical hypothesis should be rejected.

"On-line Cournot principle": events with negligible upper probabilities never happen.

Let P be a probability distribution on infinite binary sequences

Let P be a probability distribution on infinite binary sequences

null sets with respect to P

Let P be a probability distribution on infinite binary sequences

null sets with respect to P

if P is computable, one can define effectively null sets

Let P be a probability distribution on infinite binary sequences

null sets with respect to P

if P is computable, one can define effectively null sets

Maximal effectively null set; it's complement is the set of all Martin-Löf random sequences

Let P be a probability distribution on infinite binary sequences

null sets with respect to P

if P is computable, one can define effectively null sets

Maximal effectively null set; it's complement is the set of all Martin-Löf random sequences

Now let P be an on-line distribution; then the notion of on-line null set can be defined (using upper probability)

Let P be a probability distribution on infinite binary sequences

null sets with respect to P

if P is computable, one can define effectively null sets

Maximal effectively null set; it's complement is the set of all Martin-Löf random sequences

Now let P be an on-line distribution; then the notion of on-line null set can be defined (using upper probability)

If P is computable, the notion of effectively on-line null sets is defined; there exists maximal one; it's complement is the set of on-line random sequences.

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

- a priori probability coincide with P (up to a $O(1)$-factor) on its prefixes (Schnorr – Levin)

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

- a priori probability coincide with P (up to a $O(1)$-factor) on its prefixes (Schnorr – Levin)
- no lower semicomputable supermartingale is infinite on its prefixes (Schnorr);

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

- a priori probability coincide with P (up to a $O(1)$-factor) on its prefixes (Schnorr – Levin)
- no lower semicomputable supermartingale is infinite on its prefixes (Schnorr);

Now: a sequence is random with respect to an on-line probability distribution P iff

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

- a priori probability coincide with P (up to a $O(1)$-factor) on its prefixes (Schnorr – Levin)
- no lower semicomputable supermartingale is infinite on its prefixes (Schnorr);

Now: a sequence is random with respect to an on-line probability distribution P iff

- a priori on-line probability coincide with P (up to a $O(1)$-factor) on its prefixes

Criteria of randomness: a sequence is random with respect to a probability distribution P iff

- a priori probability coincide with P (up to a $O(1)$-factor) on its prefixes (Schnorr – Levin)
- no lower semicomputable supermartingale is infinite on its prefixes (Schnorr);

Now: a sequence is random with respect to an on-line probability distribution P iff

- a priori on-line probability coincide with P (up to a $O(1)$-factor) on its prefixes
- no lower semicomputable on-line supermartingale is infinite on its prefixes

Prequential randomness

Prequential randomness

A sequence $p_1, b_1, p_2, b_2, \ldots$ is given; b_i are bits, p_i are rational numbers in $(0, 1)$

Prequential randomness

A sequence $p_1, b_1, p_2, b_2, \ldots$ is given; b_i are bits, p_i are rational numbers in $(0, 1)$

somebody tells us that this sequence is a protocol of an adjustable random bit generator (b_i is obtained randomly and $b_i = 1$ with probability p_i)

Prequential randomness

A sequence $p_1, b_1, p_2, b_2, \ldots$ is given; b_i are bits, p_i are rational numbers in $(0, 1)$

somebody tells us that this sequence is a protocol of an adjustable random bit generator (b_i is obtained randomly and $b_i = 1$ with probability p_i)

sometimes we do not believe in this

Prequential randomness

A sequence $p_1, b_1, p_2, b_2, \ldots$ is given; b_i are bits, p_i are rational numbers in $(0, 1)$

somebody tells us that this sequence is a protocol of an adjustable random bit generator (b_i is obtained randomly and $b_i = 1$ with probability p_i)

sometimes we do not believe in this

e.g., all $p_i < 0.1$ and most of b_i are 1's

Prequential randomness

A sequence $p_1, b_1, p_2, b_2, \ldots$ is given; b_i are bits, p_i are rational numbers in $(0, 1)$

somebody tells us that this sequence is a protocol of an adjustable random bit generator (b_i is obtained randomly and $b_i = 1$ with probability p_i)

sometimes we do not believe in this

e.g., all $p_i < 0.1$ and most of b_i are 1's

A formal definition: we require that $p_1, b_1, p_2, b_2, \ldots$ is on-line random wrt on-line distribution where

$$\Pr[b_i = 1 | p_1, b_1, \ldots, p_i] = p_i$$

Muchnik's paradox

Let $b_1, b_2, b_3 \ldots$ is a sequence of random bits produced by two people alternatively.

Muchnik's paradox

Let $b_1, b_2, b_3 \ldots$ is a sequence of random bits produced by two people alternatively.

Each of them guarantees that her bits are random in the context of the sequence (when other's bits are external data)

Muchnik's paradox

Let $b_1, b_2, b_3 \ldots$ is a sequence of random bits produced by two people alternatively.

Each of them guarantees that her bits are random in the context of the sequence (when other's bits are external data)

Can we conclude that the entire sequence is random?

Muchnik's paradox

Let $b_1, b_2, b_3 \ldots$ is a sequence of random bits produced by two people alternatively.

Each of them guarantees that her bits are random in the context of the sequence (when other's bits are external data)

Can we conclude that the entire sequence is random?

Andrei A. Muchnik [1958–2007]: negative answer