Algorithmic independence and probabilistic arguments

Alexander Shen,
LIF CNRS & Univ. Aix – Marseille

November 2009
Joint work with
C.-L.Chang, Y.-D. Lyuu, Y.-W. Ti (Taiwan)

Algoritmic randomness

Algoritmic randomness

- Algorithmic randomness = incompressibility

Algoritmic randomness

- Algorithmic randomness $=$ incompressibility
- n-bit string x is random if $C(x) \approx n$

Algoritmic randomness

- Algorithmic randomness $=$ incompressibility
- n-bit string x is random if $C(x) \approx n$
- Randomness deficiency for n-bit string: $d(x) = n - C(x)$

Algoritmic randomness

- Algorithmic randomness = incompressibility
- n-bit string x is random if $C(x) \approx n$
- Randomness deficiency for n-bit string:
  $d(x) = n - C(x)$
- Version: $d'(x) = n - C(x|n)$;
  $|d - d'| = O(\log d)$

Algoritmic randomness

- Algorithmic randomness $=$ incompressibility
- n-bit string x is random if $C(x) \approx n$
- Randomness deficiency for n-bit string: $d(x) = n - C(x)$
- Version: $d'(x) = n - C(x|n)$; $|d - d'| = O(\log d)$
- Most strings of length n are random: the fraction of strings of deficiency $> k$ is about 2^{-k}.

Algoritmic independence

Algoritmic independence

- Two random (incompressible) strings can be dependent

Algoritmic independence

- Two random (incompressible) strings can be dependent
- General notion of independence $I(x : y) \approx 0$

Algoritmic independence

- Two random (incompressible) strings can be dependent
- General notion of independence $I(x : y) \approx 0$
- $I(x : y) = C(x) - C(x|y)$ measure how useful y is in describing x

Algoritmic independence

- Two random (incompressible) strings can be dependent
- General notion of independence $I(x:y) \approx 0$
- $I(x:y) = C(x) - C(x|y)$ measure how useful y is in describing x
- Special case of incompressible n-bit strings: $d(x|y) = n - C(x|y)$ and $d(y|x) = n - C(y|x)$ are small

Algoritmic independence

- Two random (incompressible) strings can be dependent
- General notion of independence $I(x:y) \approx 0$
- $I(x:y) = C(x) - C(x|y)$ measure how useful y is in describing x
- Special case of incompressible n-bit strings: $d(x|y) = n - C(x|y)$ and $d(y|x) = n - C(y|x)$ are small
- Another measure of dependence: $d(x,y) = 2n - C(x,y)$.

Algoritmic independence

- Two random (incompressible) strings can be dependent
- General notion of independence $I(x : y) \approx 0$
- $I(x : y) = C(x) - C(x|y)$ measure how useful y is in describing x
- Special case of incompressible n-bit strings: $d(x|y) = n - C(x|y)$ and $d(y|x) = n - C(y|x)$ are small
- Another measure of dependence: $d(x, y) = 2n - C(x, y)$.
- Relation: $d(x, y) \approx d(x) + d(y|x)$ (folklore enhancement of Kolmogorov–Levin symmetry of information)

Sets of pairwise independent strings

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ [$d(x_i, x_j) \leq D$]

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ $[d(x_i, x_j) \leq D]$
- How large N could be?

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ [$d(x_i, x_j) \leq D$]
- How large N could be?
- answer: $N = 2^{D+O(\log D)}$ (for both versions)

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ [$d(x_i, x_j) \leq D$]
- How large N could be?
- answer: $N = 2^{D+O(\log D)}$ (for both versions)
- Lower bound: choose x_i sequentially, each x_i has small $d(x_i|x_j)$ for all $j < i$.

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ $[d(x_i, x_j) \leq D]$
- How large N could be?
- answer: $N = 2^{D+O(\log D)}$ (for both versions)
- Lower bound: choose x_i sequentially, each x_i has small $d(x_i|x_j)$ for all $j < i$.
- the existence is guaranteed since each x_i disables about 2^{n-D} strings

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ [$d(x_i, x_j) \leq D$]
- How large N could be?
- answer: $N = 2^{D+O(\log D)}$ (for both versions)
- Lower bound: choose x_i sequentially, each x_i has small $d(x_i|x_j)$ for all $j < i$.
- the existence is guaranteed since each x_i disables about 2^{n-D} strings
- Problem: what about $d(x_j|x_i)$?

Sets of pairwise independent strings

- $x_1, \ldots, x_N$ are n-bit strings
- each x_i is random and each pair x_i, x_j is independent: $d(x_i|x_j) \leq D$ $[d(x_i, x_j) \leq D]$
- How large N could be?
- answer: $N = 2^{D+O(\log D)}$ (for both versions)
- Lower bound: choose x_i sequentially, each x_i has small $d(x_i|x_j)$ for all $j < i$.
- the existence is guaranteed since each x_i disables about 2^{n-D} strings
- Problem: what about $d(x_j|x_i)$?
- Solution: use only strings with randomness deficiency $O(1)$.

Sets of pairwise independent strings: upper bound

Sets of pairwise independent strings: upper bound

- If x and y have a common prefix of length m, then $d(x|y) \geq m - O(\log m)$

Sets of pairwise independent strings: upper bound

- If x and y have a common prefix of length m, then $d(x|y) \geq m - O(\log m)$
- So if $x_1, \ldots, x_N$ are pairwise independent, then their $D + O(\log D)$ bit prefixes are different

Sets of pairwise independent strings: upper bound

- If x and y have a common prefix of length m, then $d(x|y) \geq m - O(\log m)$
- So if $x_1, \ldots, x_N$ are pairwise independent, then their $D + O(\log D)$ bit prefixes are different
- so $N \leq 2^{D+O(\log D)}$.

Sets of pairwise independent strings: upper bound

- If x and y have a common prefix of length m, then $d(x|y) \geq m - O(\log m)$
- So if $x_1, \ldots, x_N$ are pairwise independent, then their $D + O(\log D)$ bit prefixes are different
- so $N \leq 2^{D+O(\log D)}$.

Remark: Lower bound is a kind of generalization of Gilbert bound in coding theory: there we look for strings that have large Hamming distance, now we require more: information distance. (Small Hamming distance $\Rightarrow$ dependence).

t-wise independence

t-wise independence

- Fix some t (say, $t = 3$).

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i | x_j, x_k) = n - C(x_i | x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)
- the second condition may be a bit stronger

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)
- the second condition may be a bit stronger
- how large N could be?

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)
- the second condition may be a bit stronger
- how large N could be?
- answer: $N = 2^{D/(t-1)+O(\log D)}$.

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)
- the second condition may be a bit stronger
- how large N could be?
- answer: $N = 2^{D/(t-1)+O(\log D)}$.
- ($t = 2$ gives the previous result)

t-wise independence

- Fix some t (say, $t = 3$).
- Looking for n-bit strings $x_1, \ldots, x_N$ such that any t of them are (almost) independent
- $d(x_i|x_j, x_k) = n - C(x_i|x_j, x_k) \leq D$ for all i, j, k (different)
- $d(x_i, x_j, x_k) = 3n - C(x_i, x_j, x_k) \leq D$ for all i, j, k (different)
- the second condition may be a bit stronger
- how large N could be?
- answer: $N = 2^{D/(t-1)+O(\log D)}$.
- ($t = 2$ gives the previous result)
- both arguments (for lower and upper bounds) do not work anymore

t-wise independence: lower bound

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$
- with positive probability they are independent?

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$
- with positive probability they are independent?
- no (let $t = 3$): for given x_i, x_j, x_k the probability of $d(x_i, x_j, x_k) > D$ is 2^{-D} and there are N^3 (not N^2) triples.

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$
- with positive probability they are independent?
- no (let $t = 3$): for given x_i, x_j, x_k the probability of $d(x_i, x_j, x_k) > D$ is 2^{-D} and there are N^3 (not N^2) triples.
- this would give $2^{D/t}$ lower bound

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$
- with positive probability they are independent?
- no (let $t = 3$): for given x_i, x_j, x_k the probability of $d(x_i, x_j, x_k) > D$ is 2^{-D} and there are N^3 (not N^2) triples.
- this would give $2^{D/t}$ lower bound
- for given i the probability of the event

$$\exists j \exists k [i, j, k \text{ are different}, d(x_i, x_j, x_k) > D]$$

is at most $1/2$.

t-wise independence: lower bound

- Take $x_1, \ldots, x_N$ randomly with $N = 2^{D/(t-1)}$
- with positive probability they are independent?
- no (let $t = 3$): for given x_i, x_j, x_k the probability of $d(x_i, x_j, x_k) > D$ is 2^{-D} and there are N^3 (not N^2) triples.
- this would give $2^{D/t}$ lower bound
- for given i the probability of the event

$$\exists j \exists k [i, j, k \text{ are different}, d(x_i, x_j, x_k) > D]$$

  is at most $1/2$.
- the expected number of bad i is at most $N/2$: delete them

t-wise independence: upper bound

t-wise independence: upper bound

- now same prefixes are not enough

t-wise independence: upper bound

- now same prefixes are not enough
- but we have three strings instead of two; if x_i, x_j, x_k have prefixes p_i, p_j, p_k such that $\varphi(p_i, p_j) = p_k$ for some simple function, this is enough for us

t-wise independence: upper bound

- now same prefixes are not enough
- but we have three strings instead of two; if x_i, x_j, x_k have prefixes p_i, p_j, p_k such that $\varphi(p_i, p_j) = p_k$ for some simple function, this is enough for us
- combinatorial question: given a set A, find a function $\varphi\colon A \times A \to A$ such that for all sufficiently large $B \subset A$ there exist three different elements $b_1, b_2, b_3 \in B$ such that $\varphi(b_1, b_2) = b_3$

t-wise independence: upper bound

- now same prefixes are not enough
- but we have three strings instead of two; if x_i, x_j, x_k have prefixes p_i, p_j, p_k such that $\varphi(p_i, p_j) = p_k$ for some simple function, this is enough for us
- combinatorial question: given a set A, find a function $\varphi\colon A \times A \to A$ such that for all sufficiently large $B \subset A$ there exist three different elements $b_1, b_2, b_3 \in B$ such that $\varphi(b_1, b_2) = b_3$
- "sufficiently large": a bit larger that $\sqrt{\#A}$

t-wise independence: upper bound

- now same prefixes are not enough
- but we have three strings instead of two; if x_i, x_j, x_k have prefixes p_i, p_j, p_k such that $\varphi(p_i, p_j) = p_k$ for some simple function, this is enough for us
- combinatorial question: given a set A, find a function $\varphi\colon A \times A \to A$ such that for all sufficiently large $B \subset A$ there exist three different elements $b_1, b_2, b_3 \in B$ such that $\varphi(b_1, b_2) = b_3$
- "sufficiently large": a bit larger that $\sqrt{\#A}$
- existence: probabilistic argument

t-wise independence: upper bound

- now same prefixes are not enough
- but we have three strings instead of two; if x_i, x_j, x_k have prefixes p_i, p_j, p_k such that $\varphi(p_i, p_j) = p_k$ for some simple function, this is enough for us
- combinatorial question: given a set A, find a function $\varphi\colon A \times A \to A$ such that for all sufficiently large $B \subset A$ there exist three different elements $b_1, b_2, b_3 \in B$ such that $\varphi(b_1, b_2) = b_3$
- "sufficiently large": a bit larger that $\sqrt{\#A}$
- existence: probabilistic argument
- existence of a simple φ: exhaustive search