Topic Course on Probabilistic Methods

(Week 4)

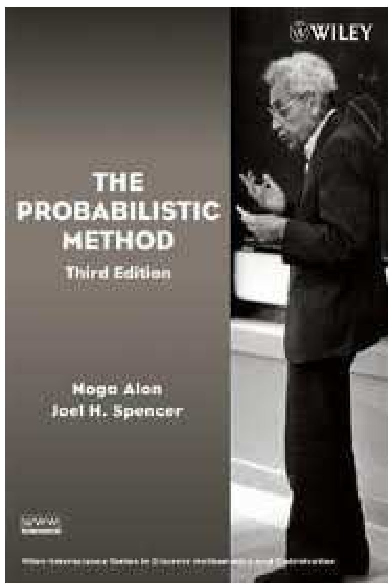
Second Moment Method

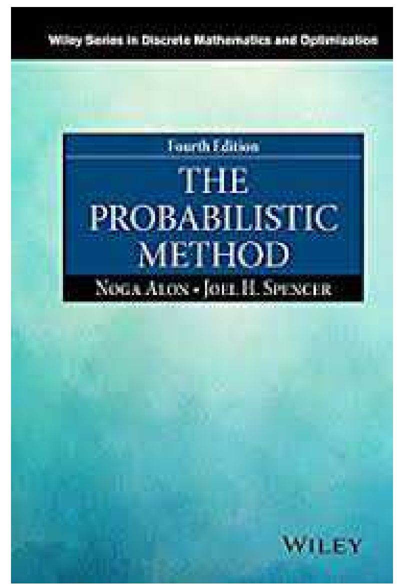
Linyuan Lu

University of South Carolina

Introduction

The topic course is mostly based the textbook "The probabilistic Method" by Noga Alon and Joel Spencer (third edition 2008, John Wiley & Sons, Inc. ISBN 9780470170205 or fourth edition ISBN-13: 978-1119061953.)





Selected topics

- Linearity of Expectation (2 weeks)
- Alterations (1 week)
- The second moment method (1 week)
- The Local Lemma (1-2 weeks)
- Correlation Inequalities (1 week)
- Large deviations (1-2 weeks)
- Poisson Paradigm (1 week)
- Random graphs (2 weeks)
- Discrepancy (1 week)
- Entropy (1 week)

Subtopics

The second moment method

- Variance
- Chebyshev's inequality
- Number of prime factors
- Counting K_4 in $G(n, p)$
- Counting balanced graphs
- Clique number of $G(n, \frac{1}{2})$
- Distinct sum

Variance

- Variance:
  $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$

Variance

- Variance:
  $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.
- Co-Variance: $\mathrm{Cov}(X, Y) =$
  $\mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$

Variance

- Variance:
  $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.
- Co-Variance: $\mathrm{Cov}(X, Y) =$
  $\mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

Variance

- Variance:
  $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$
- Co-Variance: $\mathrm{Cov}(X, Y) =$
  $\mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0.$

If $X = \sum_{i=1}^{n} X_i$, then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

Variance

- Variance:
  $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$
- Co-Variance: $\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0.$

If $X = \sum_{i=1}^n X_i$, then

$$\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

If $X_1, \ldots, X_n$ are mutually independent, then
$\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i).$

Chebyshev's Inequality

- $\mathrm{E}(X) = \mu$,
- $\mathrm{Var}(X) = \sigma^2$.

Theorem [Chebyshev's Inequality]: For any positive λ,

$$\Pr(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Chebyshev's Inequality

- $\mathrm{E}(X) = \mu$,
- $\mathrm{Var}(X) = \sigma^2$.

Theorem [Chebyshev's Inequality]: For any positive λ,

$$\Pr(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Proof:

$$\begin{aligned} \sigma^2 &= \mathrm{Var}(X) \\ &= \mathrm{E}((X-\mu)^2) \\ &\geq \lambda^2\sigma^2 \Pr(|X - \mu| \geq \lambda\sigma). \end{aligned}$$

Number theory

$\nu(n)$: the number of primes p dividing n.

Number theory

$\nu(n)$: the number of primes p dividing n.

Hardy, Ramanujan [1920]: For "almost all" n, $\nu(n) \approx \ln \ln n$.

Number theory

$\nu(n)$: the number of primes p dividing n.

Hardy, Ramanujan [1920]: For "almost all" n, $\nu(n) \approx \ln \ln n$.

Theorem [Turán (1934)]: Let $\omega(n) \to \infty$ arbitrarily slowly. Then the number of x in $[n] := \{1, 2, \ldots, n\}$ such that

$$|\nu(x) - \ln \ln n| > \omega(n)\sqrt{\ln \ln n}.$$

is $o(n)$.

Proof

Let x be randomly chosen from $[n]$. For p prime set

$$X_p = \begin{cases} 1 & \text{if } p \mid x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$ and $X = \sum_{p \le M} X_p$. Then

$$\nu(x) - 10 \le X(x) \le \nu(x).$$

Proof

Let x be randomly chosen from $[n]$. For p prime set

$$X_p = \begin{cases} 1 & \text{if } p \mid x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$ and $X = \sum_{p \leq M} X_p$. Then

$$\nu(x) - 10 \leq X(x) \leq \nu(x).$$

$$\mathrm{E}(X_p) = \frac{\lfloor \frac{n}{p} \rfloor}{n} = \frac{1}{p} + O(\frac{1}{n}).$$

Proof

Let x be randomly chosen from $[n]$. For p prime set

$$X_p = \begin{cases} 1 & \text{if } p \mid x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$ and $X = \sum_{p \leq M} X_p$. Then

$$\nu(x) - 10 \leq X(x) \leq \nu(x).$$

$$\mathrm{E}(X_p) = \frac{\lfloor \frac{n}{p} \rfloor}{n} = \frac{1}{p} + O(\frac{1}{n}).$$

$$\mathrm{E}(X) = \sum_{p \leq M} \left(\frac{1}{p} + O(\frac{1}{n}) \right) = \ln \ln n + O(1).$$

Variance of $\mathrm{E}(X)$

$$\mathrm{Var}(X_p) = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{n}\right).$$

Variance of $\mathrm{E}(X)$

$$\mathrm{Var}(X_p) = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{n}\right).$$

$$\begin{aligned}
\mathrm{Cov}(X_p, X_q) &= \mathrm{E}(X_p X_q) - \mathrm{E}(X_p)\mathrm{E}(X_q) \\
&= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\
&\leq \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \\
&\leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).
\end{aligned}$$

Variance of $\mathrm{E}(X)$

$$\mathrm{Var}(X_p) = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{n}\right).$$

$$\begin{aligned}
\mathrm{Cov}(X_p, X_q) &= \mathrm{E}(X_p X_q) - \mathrm{E}(X_p)\mathrm{E}(X_q) \\
&= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\
&\le \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \\
&\le \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).
\end{aligned}$$

$$\mathrm{Var}(X) = \sum_{p \le M} \mathrm{Var}(X_p) + \sum_{p \ne q} \mathrm{Cov}(X_p, X_q) = \ln\ln n + O(1).$$

continue

By Chebyshev's inequality, we have

$$\Pr(|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}) < \lambda^{-2} + o(1).$$

$\square$

continue

By Chebyshev's inequality, we have

$$\Pr(|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}) < \lambda^{-2} + o(1).$$

□

Theorem [Erdős-Kac (1940):] For any fixed λ, we have

$$\lim_{n\to\infty} \frac{1}{n}\left|\{x\colon 1 \le x \le n, \nu(x) \ge \ln\ln n + \lambda\sqrt{\ln\ln n}\}\right|$$
$$= \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Basic facts

- $X = \sum_i X_i$: where X_i are indicator random variables.

Basic facts

- $X = \sum_i X_i$: where X_i are indicator random variables.
- If $\mathrm{E}(X) = o(1)$, then $X = 0$ almost always.

Basic facts

- $X = \sum_i X_i$: where X_i are indicator random variables.
- If $\mathrm{E}(X) = o(1)$, then $X = 0$ almost always.
- If $\mathrm{Var}(X) = o(\mathrm{E}(X)^2)$, then $X \sim \mathrm{E}(X)$ almost always. In particular $X > 0$ almost always.

Basic facts

- $X = \sum_i X_i$: where X_i are indicator random variables.
- If $\mathrm{E}(X) = o(1)$, then $X = 0$ almost always.
- If $\mathrm{Var}(X) = o(\mathrm{E}(X)^2)$, then $X \sim \mathrm{E}(X)$ almost always. In particular $X > 0$ almost always.
- Write $X_i \sim X_j$ if $i \neq j$, and the events X_i, X_j are not independent. Let $\Delta = \sum_{i \sim j} \mathrm{Pr}(A_i \wedge A_j)$. If $E(X) \to \infty$ and $\Delta = o(E(X)^2)$, then $X > 0$ almost always.

Basic facts

- $X = \sum_i X_i$: where X_i are indicator random variables.
- If $\mathrm{E}(X) = o(1)$, then $X = 0$ almost always.
- If $\mathrm{Var}(X) = o(\mathrm{E}(X)^2)$, then $X \sim \mathrm{E}(X)$ almost always. In particular $X > 0$ almost always.
- Write $X_i \sim X_j$ if $i \neq j$, and the events X_i, X_j are not independent. Let $\Delta = \sum_{i \sim j} \Pr(A_i \wedge A_j)$. If $E(X) \to \infty$ and $\Delta = o(E(X)^2)$, then $X > 0$ almost always.
- Let $\Delta^* = \max_i \sum_{j \sim i} \Pr(A_j | A_i)$. If $E(X) \to \infty$ and $\Delta^* = o(E(X))$, then $X > 0$ almost always.

Erdős-Rényi model $G(n, p)$

- n nodes

Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.

Erdős-Rényi model $G(n, p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.

Erdős-Rényi model $G(n,p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.

A property of graphs is a family of graphs closed under isomorphic.

Erdős-Rényi model $G(n,p)$

- n nodes
- For each pair of vertices, create an edge independently with probability p.
- The graph with e edges has the probability $p^e(1-p)^{\binom{n}{2}-e}$.

A property of graphs is a family of graphs closed under isomorphic.

A function $r(n)$ is called a threshold function for some property P if

- If $p \ll r(n)$, then $G(n,p)$ does not satisfy P almost always.
- If $p \gg r(n)$, then $G(n,p)$ satisfy P almost always.

Threshold of $\omega(G) \geq 4$

$\omega(G)$: the number of vertices in the maximum clique of the graph G.

Threshold of $\omega(G) \geq 4$

$\omega(G)$: the number of vertices in the maximum clique of the graph G.

Theorem: The property $\omega(G) \geq 4$ has the threshold function $n^{-2/3}$.

Threshold of $\omega(G) \geq 4$

$\omega(G)$: the number of vertices in the maximum clique of the graph G.

Theorem: The property $\omega(G) \geq 4$ has the threshold function $n^{-2/3}$.

Proof: For any $S \in \binom{[n]}{4}$, let X_S be the indicator variable of the event "S is a clique".

Threshold of $\omega(G) \geq 4$

$\omega(G)$: the number of vertices in the maximum clique of the graph G.

Theorem: The property $\omega(G) \geq 4$ has the threshold function $n^{-2/3}$.

Proof: For any $S \in \binom{[n]}{4}$, let X_S be the indicator variable of the event "S is a clique".

$$\mathrm{E}(X) = \sum_S \mathrm{E}(X_S) = \binom{n}{4} p^6 \approx \frac{n^4 p^6}{24}.$$

Threshold of $\omega(G) \geq 4$

$\omega(G)$: the number of vertices in the maximum clique of the graph G.

Theorem: The property $\omega(G) \geq 4$ has the threshold function $n^{-2/3}$.

Proof: For any $S \in \binom{[n]}{4}$, let X_S be the indicator variable of the event "S is a clique".

$$\mathrm{E}(X) = \sum_S \mathrm{E}(X_S) = \binom{n}{4} p^6 \approx \frac{n^4 p^6}{24}.$$

If $p \ll n^{-2/3}$ then $\mathrm{E}(X) = o(1)$ and so $X = 0$ almost surely.

Continue

If $p \gg n^{-2/3}$, then $\mathrm{E}(X) \to \infty$.

Continue

If $p \gg n^{-2/3}$, then $\mathrm{E}(X) \to \infty$.

$S \sim T$ if $|S \cap T| \geq 2$. Thus,

$$\Delta^* = O(n^2p^5) + O(np^3) = o(n^4p^6) = o(\mathrm{E}(X)).$$

Hence $X > 0$ almost surely. □

Balanced graphs

- H has v vertices and e edges.

Balanced graphs

- H has v vertices and e edges.
- $\rho(H) = e/v$.

Balanced graphs

- H has v vertices and e edges.
- $\rho(H) = e/v$.
- H is called **balanced** of for any subgraph H',

$$\rho(H') \leq \rho(H).$$

Balanced graphs

- H has v vertices and e edges.
- $\rho(H) = e/v$.
- H is called **balanced** of for any subgraph H',

$$\rho(H') \leq \rho(H).$$

- H is called **strictly balanced** of for any proper subgraph H',

$$\rho(H') < \rho(H).$$

Results

Theorem: *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

Results

Theorem: *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

If H is not balanced then $p = n^{-v/e}$ is the threshed function for A.

Results

Theorem: *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

If H is not balanced then $p = n^{-v/e}$ is the threshed function for A.

Proof: Write $X = \sum_S X_S$. Then $\mathrm{E}(X) = \binom{n}{v} p^e$.

Results

Theorem: *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

If H is not balanced then $p = n^{-v/e}$ is the threshed function for A.

Proof: Write $X = \sum_S X_S$. Then $\mathrm{E}(X) = \binom{n}{v} p^e$.

If $p \ll n^{-v/e}$, then $\mathrm{E}(X) = o(1)$; $X = 0$ almost surely.

Results

Theorem: *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

If H is not balanced then $p = n^{-v/e}$ is the threshed function for A.

Proof: Write $X = \sum_S X_S$. Then $\mathrm{E}(X) = \binom{n}{v} p^e$.

If $p \ll n^{-v/e}$, then $\mathrm{E}(X) = o(1)$; $X = 0$ almost surely.

If $p \gg n^{-v/e}$, then $\mathrm{E}(X) \to \infty$. We have

$$\Delta^* = O(\sum_{i=2}^{v} n^{v-i} p^{e-(ie/v)}) = o(\mathrm{E}(X)).$$

Two other results

Theorem: *Let H be a strictly balanced graph with v vertices and e edges and a automorphisms. Let X be the copies of H in $G(n,p)$. Assume $p \gg n^{-v/s}$. Then almost always*

$$X \sim \frac{n^v p^e}{a}.$$

Two other results

Theorem: *Let H be a strictly balanced graph with v vertices and e edges and a automorphisms. Let X be the copies of H in $G(n,p)$. Assume $p \gg n^{-v/s}$. Then almost always*

$$X \sim \frac{n^v p^e}{a}.$$

Theorem: *Let H be any fixed graph. For every subgraph H' of H (including H itself) let $X_{H'}$ denote the number of copies of H' in $G(n,p)$. Assume p is such that $E(X_{H'}) \to \infty$ for every H'. Then almost surely*

$$X_H \sim \mathrm{E}(X_H).$$

Clique number of $G(n, 1/2)$

- $\omega(G)$: the clique number of G.

Clique number of $G(n, 1/2)$

- $\omega(G)$: the clique number of G.
- $f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$: the expected number of k-cliques.

Clique number of $G(n, 1/2)$

- $\omega(G)$: the clique number of G.
- $f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$: the expected number of k-cliques.

Theorem: Let $k = k(n)$ satisfying $k \sim 2 \log_2 n$ and $f(k) \to \infty$. Then almost surely $\omega(G) \geq k$.

Clique number of $G(n, 1/2)$

- $\omega(G)$: the clique number of G.
- $f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$: the expected number of k-cliques.

Theorem: Let $k = k(n)$ satisfying $k \sim 2\log_2 n$ and $f(k) \to \infty$. Then almost surely $\omega(G) \geq k$.

Proof: For each k-set S, let X_S be the indicator random variable that S is a clique and $X = \sum_{|S|=k} X_S$.

$$\mathrm{E}(X) = \binom{n}{k} 2^{-\binom{k}{2}} = f(k).$$

Continue

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}}.$$

$$\frac{\Delta^*}{E(|X|)} = \sum_{i=2}^{k-1} g(i),$$

where $g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}$.

Continue

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}-\binom{k}{2}}.$$

$$\frac{\Delta^*}{E(|X|)} = \sum_{i=2}^{k-1} g(i),$$

where $g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}$. Then

$$g(i) \leq \max\{g(2), g(k-1)\} = o(n^{-1}).$$

Thus, $\Delta^* = o(\mathrm{E}(X))$. $\square$

Remark

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1}2^{-k}.$$

For $k \sim 2\log_2 n$, then

$$\frac{f(k+1)}{f(k)} = n^{-1+o(1)}.$$

Remark

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1} 2^{-k}.$$

For $k \sim 2\log_2 n$, then

$$\frac{f(k+1)}{f(k)} = n^{-1+o(1)}.$$

Let k_0 be the value with $f(k_0) \geq 1 > f(k_0+1)$. For most of n, $f(k)$ will jump from very large to very small. With high probability, $\omega(G) = k_0$.

Distinct sums

- A set $x_1, \ldots, x_k$ of positive integers is said to have **distinct sums** if all sums

$$\sum_{i \in S} x_i, \quad S \subset \{1, \ldots, k\}$$

are distinct.

Distinct sums

- A set $x_1, \ldots, x_k$ of positive integers is said to have **distinct sums** if all sums

$$\sum_{i \in S} x_i, \quad S \subset \{1, \ldots, k\}$$

are distinct.

- Let $f(n)$ be the largest k for which there is a set $\{x_1, x_2, \ldots, x_k\} \subset \{1, \ldots, n\}$ with distinct sums.

Distinct sums

- A set $x_1, \ldots, x_k$ of positive integers is said to have **distinct sums** if all sums

$$\sum_{i \in S} x_i, \quad S \subset \{1, \ldots, k\}$$

are distinct.

- Let $f(n)$ be the largest k for which there is a set $\{x_1, x_2, \ldots, x_k\} \subset \{1, \ldots, n\}$ with distinct sums.

Erdős offered $300 for a proof or disproof that

$$f(n) \leq \log_2 n + O(1).$$

An easy lower bound

Proposition: $f(n) \geq 1 + \lfloor \log_2 n \rfloor$.

An easy lower bound

Proposition: $f(n) \geq 1 + \lfloor \log_2 n \rfloor$.

Proof: Let $k = \lfloor \log_2 n \rfloor + 1$. For $i = 1, 2, \ldots, k$, set $x_i = 2^{i-1}$. We have

$$x_1 < x_2 < \cdots < x_k \leq n.$$

An easy lower bound

Proposition: $f(n) \geq 1 + \lfloor \log_2 n \rfloor$.

Proof: Let $k = \lfloor \log_2 n \rfloor + 1$. For $i = 1, 2, \ldots, k$, set $x_i = 2^{i-1}$. We have

$$x_1 < x_2 < \cdots < x_k \leq n.$$

For $\epsilon_1, \ldots, \epsilon_k \in \{0, 1\}$, the binary number

$$\epsilon_k \epsilon_{k-1} \cdots \epsilon_1$$

has the value $\sum_{i=1}^{k} \epsilon_i x_i$.

An easy lower bound

Proposition: $f(n) \geq 1 + \lfloor \log_2 n \rfloor$.

Proof: Let $k = \lfloor \log_2 n \rfloor + 1$. For $i = 1, 2, \ldots, k$, set $x_i = 2^{i-1}$. We have

$$x_1 < x_2 < \cdots < x_k \leq n.$$

For $\epsilon_1, \ldots, \epsilon_k \in \{0, 1\}$, the binary number

$$\epsilon_k \epsilon_{k-1} \cdots \epsilon_1$$

has the value $\sum_{i=1}^{k} \epsilon_i x_i$.

Since every number has a unique base-2 representation, the set $\{x_1, x_2, \ldots, x_k\}$ has distinct sums. □

An easy upper bound

Proposition: $f(n) \le \log_2 n + \log_2 \log_2 n + O(1)$.

An easy upper bound

Proposition: $f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$.

Proof: Assume $n \geq 2$. Let $k = f(n)$ and $x_1, x_2, \ldots, x_k$ be the set of distinct sums.

An easy upper bound

Proposition: $f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$.

Proof: Assume $n \geq 2$. Let $k = f(n)$ and $x_1, x_2, \ldots, x_k$ be the set of distinct sums.

All 2^k sums are distinct and less than nk. We have

$$2^k \leq nk.$$

An easy upper bound

Proposition: $f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$.

Proof: Assume $n \geq 2$. Let $k = f(n)$ and $x_1, x_2, \ldots, x_k$ be the set of distinct sums.

All 2^k sums are distinct and less than nk. We have

$$2^k \leq nk.$$

The function $g(k) := \frac{2^k}{k}$ is an increasing function. At $k = \log_2 n + \log_2 \log_2 n + 2$, we have

$$\frac{2^k}{k} = \frac{4n \log_2 n}{\log_2 n + \log_2 \log_2 n + 2} > n.$$

Thus, $k \leq \log_2 n + \log_2 \log_2 n + 2$. □

A nontrivial result

Theorem: $f(n) < \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1).$

A nontrivial result

Theorem: $f(n) < \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1)$.

Proof: Fix $x_1, x_2, \ldots, x_k$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be uniform independent $\{0,1\}$-random variables. Let $X = \sum_{i=1}^{k} \epsilon_i x_i$. We have

A nontrivial result

Theorem: $f(n) < \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$.

Proof: Fix $x_1, x_2, \ldots, x_k$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be uniform independent $\{0, 1\}$-random variables. Let $X = \sum_{i=1}^{k} \epsilon_i x_i$. We have

- $\mathrm{E}(X) = \frac{1}{2}\sum_{i=1}^{k} x_i$.

A nontrivial result

Theorem: $f(n) < \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1)$.

Proof: Fix $x_1, x_2, \ldots, x_k$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be uniform independent $\{0, 1\}$-random variables. Let $X = \sum_{i=1}^k \epsilon_i x_i$. We have

- $\mathrm{E}(X) = \frac{1}{2} \sum_{i=1}^k x_i$.
- $\mathrm{Var}(X) = \frac{1}{4} \sum_{i=1}^k x_i^2 \leq \frac{n^2 k}{4}$.

A nontrivial result

Theorem: $f(n) < \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$.

Proof: Fix $x_1, x_2, \ldots, x_k$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be uniform independent $\{0,1\}$-random variables. Let $X = \sum_{i=1}^k \epsilon_i x_i$. We have

- $\mathrm{E}(X) = \frac{1}{2}\sum_{i=1}^k x_i$.
- $\mathrm{Var}(X) = \frac{1}{4}\sum_{i=1}^k x_i^2 \leq \frac{n^2k}{4}$.

By Chebyshev's inequality, for any $\lambda > 1$,

$$\Pr\left[|X-\mu| \geq \lambda n\sqrt{k}/2\right] \leq \frac{1}{\lambda^2}.$$

continue

By the property of distinct sums, $\Pr(X = x)$ is either 0 or 2^{-k} for any integer x. Thus,

$$\Pr\left[|X-\mu| < \lambda n\sqrt{k}/2\right] \leq \frac{\lambda n\sqrt{k}+1}{2^k}.$$

continue

By the property of distinct sums, $\Pr(X = x)$ is either 0 or 2^{-k} for any integer x. Thus,

$$\Pr\left[|X - \mu| < \lambda n\sqrt{k}/2\right] \leq \frac{\lambda n\sqrt{k} + 1}{2^k}.$$

Combining with previous Chebyshev's inequality, we get

$$1 \leq \frac{1}{\lambda^2} + \frac{\lambda n\sqrt{k} + 1}{2^k}.$$

Choose $\lambda = \sqrt{3}$ and $k = \log_2 n + \frac{1}{2}\log_2\log_2 n + C$ with sufficiently large constant C. The above inequality is not satisfied. Contradiction! □