PROBABILISTIC ANALYSIS OF AN EXHAUSTIVE SEARCH ALGORITHM IN RANDOM GRAPHS

Hsien-Kuei Hwang

Academia Sinica, Taiwan
(joint work with Cyril Banderier, Vlady Ravelomanana, Vytas Zacharovas)

AofA 2008, Maresias, Brazil
April 14, 2008
An independent (or stable) set in a graph is a set of vertices no two of which share the same edge.

Maximum independent set (MIS)

The MIS problem asks for an independent set with the largest size.

\[
\text{MIS} = \{1, 3, 5, 7\}
\]
Independent set

An independent (or stable) set in a graph is a set of vertices no two of which share the same edge.

\[
\text{MIS} = \{1, 3, 5, 7\}
\]

Maximum independent set (MIS)

The MIS problem asks for an independent set with the largest size.

NP hard!!
Independent set

An independent (or stable) set in a graph is a set of vertices no two of which share the same edge.

Maximum independent set (MIS)

The MIS problem asks for an independent set with the largest size.

\[ \text{MIS} = \{1, 3, 5, 7\} \]
Equivalent versions

The same problem as \textbf{MAXIMUM CLIQUE} on the complementary graph (clique = complete subgraph).

Since the complement of a vertex cover in any graph is an independent set, MIS is equivalent to \textbf{MINIMUM VERTEX COVERING}. (A vertex cover is a set of vertices where every edge connects at least one vertex.)

Among Karp’s (1972) original list of 21 NP-complete problems.
Equivalent versions

The same problem as MAXIMUM CLIQUE on the complementary graph (clique = complete subgraph).

Since the complement of a vertex cover in any graph is an independent set, MIS is equivalent to MINIMUM VERTEX COVERING. (A vertex cover is a set of vertices where every edge connects at least one vertex.)

Among Karp’s (1972) original list of 21 NP-complete problems.
Equivalent versions

The same problem as MAXIMUM CLIQUE on the complementary graph (clique = complete subgraph).

Since the complement of a vertex cover in any graph is an independent set, MIS is equivalent to MINIMUM VERTEX COVERING. (A vertex cover is a set of vertices where every edge connects at least one vertex.)

Among Karp’s (1972) original list of 21 NP-complete problems.
Random models: Erdős-Rényi’s $G_{n,p}$

Vertex set = \{1, 2, \ldots, n\} and all edges occur independently with the same probability $p$.

The cardinality of an MIS in $G_{n,p}$

Matula (1970), Grimmett and McDiarmid (1975), Bollobas and Erdős (1976), Frieze (1990): If $pn \to \infty$, then $(q := 1 - p)$

$$|\text{MIS}_n| \sim 2 \log_{1/q} pn \quad \text{whp},$$

where $q = 1 - p$; and $\exists k = k_n$ such that

$$|\text{MIS}_n| = k \text{ or } k + 1 \quad \text{whp}.$$
THEORETICAL RESULTS

Random models: Erdős-Rényi’s $G_{n,p}$

Vertex set = $\{1, 2, \ldots, n\}$ and all edges occur independently with the same probability $p$.

The cardinality of an MIS in $G_{n,p}$

Matula (1970), Grimmett and McDiarmid (1975), Bollobas and Erdős (1976), Frieze (1990): If $pn \to \infty$, then ($q := 1 - p$)

$$|\text{MIS}_n| \sim 2 \log_{1/q} pn \text{ whp,}$$

where $q = 1 - p$; and $\exists k = k_n$ such that

$$|\text{MIS}_n| = k \text{ or } k + 1 \text{ whp.}$$
Adding vertices one after another whenever possible

**The size of the resulting IS:**

\[ S_n \overset{d}{=} 1 + S_{n-1} - \text{Binom}(n-1; p) \quad (n \geq 1), \]

with \( S_0 \equiv 0 \).

Equivalent to the length of the right arm of random digital search trees.
A GREEDY ALGORITHM

Adding vertices one after another whenever possible

The size of the resulting IS:

\[ S_n \overset{d}{=} 1 + S_{n-1} - \text{Binom}(n-1; p) \]  
\( (n \geq 1), \)

with \( S_0 = 0. \)

Equivalent to the length of the right arm of random digital search trees.
Easy for people in this community

- **Mean:** \( \mathbb{E}(S_n) \sim \log_{1/q} n + \text{a bounded periodic function.} \)

- **Variance:** \( \mathbb{V}(S_n) \sim \text{a bounded periodic function.} \)

- **Limit distribution does not exist:**
  \[
  \mathbb{E} \left( e^{(X_n - \log_{1/q} n)y} \right) \sim F(\log_{1/q} n; y), \text{ where }
  \]
  \[
  F(u; y) := \frac{1 - e^y}{\log(1/q)} \left( \prod_{\ell \geq 1} \frac{1 - e^y q^\ell}{1 - q^\ell} \right) \sum_{j \in \mathbb{Z}} \Gamma \left( -\frac{y + 2j \pi i}{\log(1/q)} \right) e^{2j \pi i u}.
  \]
Easy for people in this community

- **Mean:** $\mathbb{E}(S_n) \sim \log_{1/q} n + \text{a bounded periodic function.}$

- **Variance:** $\mathbb{V}(S_n) \sim \text{a bounded periodic function.}$

- **Limit distribution does not exist:**

  $\mathbb{E}\left(e^{(X_n - \log_{1/q} n)y}\right) \sim F(\log_{1/q} n; y), \text{ where}$

  $$F(u; y) := \frac{1 - e^y}{\log(1/q)} \left(\prod_{\ell \geq 1} \frac{1 - e^y q^\ell}{1 - q^\ell}\right) \sum_{j \in \mathbb{Z}} \Gamma \left(-\frac{y + 2j\pi i}{\log(1/q)}\right) e^{2j\pi iu}.$$
ANALYSIS OF THE GREEDY ALGORITHM

Easy for people in this community

- **Mean:** $\mathbb{E}(S_n) \sim \log_{1/q} n + \text{a bounded periodic function}. $

- **Variance:** $\mathbb{V}(S_n) \sim \text{a bounded periodic function}. $

- **Limit distribution does not exist:**

$$
\mathbb{E} \left( e^{(X_n - \log_{1/q} n) y} \right) \sim F(\log_{1/q} n; y), \text{ where }
$$

$$
F(u; y) := \frac{1 - e^y}{\log(1/q)} \left( \prod_{\ell \geq 1} \frac{1 - e^y q^\ell}{1 - q^\ell} \right) \sum_{j \in \mathbb{Z}} \Gamma \left( -\frac{y + 2j\pi i}{\log(1/q)} \right) e^{2j\pi i u}.
$$
A BETTER ALGORITHM?

Goodness of GREEDY IS


Asymptotically, the GREEDY IS is half optimal.

Can we do better?

Frieze and McDiarmid (1997, RSA), Algorithmic theory of random graphs, Research Problem 15:

Construct a polynomial time algorithm that finds an independent set of size at least \((1/2 + \varepsilon)|MIS_n|\) whp or show that such an algorithm does not exist modulo some reasonable conjecture in the theory of computational complexity such as, e.g., \(P \neq NP\).


**Goodness of GREEDY IS**


Asymptotically, the GREEDY IS is half optimal.

**Can we do better?**

Frieze and McDiarmid (1997, RSA), Algorithmic theory of random graphs, Research Problem 15:

*Construct a polynomial time algorithm that finds an independent set of size at least* \( \left( \frac{1}{2} + \varepsilon \right) |MIS_n| \) *whp or show that such an algorithm does not exist modulo some reasonable conjecture in the theory of computational complexity such as, e.g., \( P \neq NP \).*
A BETTER ALGORITHM?

Goodness of GREEDY IS


Asymptotically, the GREEDY IS is half optimal.

Can we do better?

Frieze and McDiarmid (1997, RSA), Algorithmic theory of random graphs, Research Problem 15:

Construct a polynomial time algorithm that finds an independent set of size at least \( \left( \frac{1}{2} + \varepsilon \right) |MIS_n| \) whp or show that such an algorithm does not exist modulo some reasonable conjecture in the theory of computational complexity such as, e.g., \( P \neq NP \).
A degenerate form of simulated annealing

Sequentially increase the clique ($K$) size by: (i) choose a vertex $v$ u.a.r. from $V$; (ii) if $v \not\in K$ and $v$ connected to every vertex of $K$, then add $v$ to $K$; (iii) if $v \in K$, then $v$ is subtracted from $K$ with probability $\lambda^{-1}$.

He showed: $\forall \lambda \geq 1, \exists$ an initial state from which the expected time for the Metropolis process to reach a clique of size at least $(1 + \varepsilon) \log_{1/q}(pn)$ exceeds $n^{\Omega(\log pn)}$.

$n^\log n = e^{(\log n)^2}$
A degenerate form of simulated annealing

Sequentially increase the clique \((K)\) size by: (i) choose a vertex \(v\) u.a.r. from \(V\); (ii) if \(v \notin K\) and \(v\) connected to every vertex of \(K\), then add \(v\) to \(K\); (iii) if \(v \in K\), then \(v\) is subtracted from \(K\) with probability \(\lambda^{-1}\).

He showed: \(\forall \lambda \geq 1, \exists\) an initial state from which the expected time for the Metropolis process to reach a clique of size at least \((1 + \varepsilon) \log_{1/q} (pn)\) exceeds \(n^{\Omega(\log pn)}\).

\[ n^{\log n} = e^{(\log n)^2} \]
Exact algorithms

A huge number of algorithms proposed in the literature; see Bomze et al.’s survey (in *Handbook of Combinatorial Optimization*, 1999).

Special algorithms

- Wilf’s (1986) *Algorithms and Complexity* describes a *backtracking* algorithm enumerating all independent sets with time complexity $n^{O(\log n)}$.
- Chvátal (1977) proposes *exhaustive* algorithms where almost all $G_{n,1/2}$ creates at most $n^2(1+\log_2 n)$ subproblems.
- Pittel (1982):

$$P \left( n^{1-\epsilon} \log_{1/\delta} n < \text{Time used by Chvátal's algo} < n^{1+\epsilon} \log_{1/\delta} n \right) \geq 1 - e^{-c \log^2 n}$$
Exact algorithms

A huge number of algorithms proposed in the literature; see Bomze et al.’s survey (in Handbook of Combinatorial Optimization, 1999).

Special algorithms

- Wilf’s (1986) *Algorithms and Complexity* describes a *backtracking* algorithm enumerating all independent sets with time complexity \( n^{O(\log n)} \).

- Chvátal (1977) proposes *exhaustive* algorithms where almost all \( G_{n,1/2} \) creates at most \( n^{2(1+\log_2 n)} \) subproblems.

- Pittel (1982):

\[
\mathbb{P} \left( n^{\frac{1-\varepsilon}{4} \log \frac{1}{q} n} \leq \text{Time used by Chvátal’s algo} \leq n^{\frac{1+\varepsilon}{2} \log \frac{1}{q} n} \right) \geq 1 - e^{-c \log^2 n}
\]
Exact algorithms

A huge number of algorithms proposed in the literature; see Bomze et al.’s survey (in *Handbook of Combinatorial Optimization*, 1999).

Special algorithms

- Wilf’s (1986) *Algorithms and Complexity* describes a *backtracking* algorithm enumerating all independent sets with time complexity $n^{O(\log n)}$.

- Chvátal (1977) proposes *exhaustive* algorithms where almost all $G_{n,1/2}$ creates at most $n^{2(1+\log_2 n)}$ subproblems.

- Pittel (1982):

$$\mathbb{P}\left(n^{\frac{1-\varepsilon}{4}} \log_{1/q} n \leq \text{Time used by Chvátal’s algo} \leq n^{\frac{1+\varepsilon}{2}} \log_{1/q} n\right) \geq 1 - e^{-c \log^2 n}$$
MIS contains either \( v \) or not

\[
X_n \triangleq X_{n-1} + X_{n-1}^* - \text{Binom}(n-1; p) \quad (n \geq 2),
\]

with \( X_0 = 0 \) and \( X_1 = 1 \).

**Special cases**

- If \( p \) is close to 1, then the second term is small, so we expect a *polynomial* time bound.
- If \( p \) is sufficiently small, then the second term is large, and we expect an *exponential* time bound.
- What happens for \( p \) in between?
MIS contains either \( v \) or not

\[
X_n \overset{d}{=} X_{n-1} + X_{n-1}^* - \text{Binom}(n-1; p) \quad (n \geq 2),
\]

with \( X_0 = 0 \) and \( X_1 = 1 \).

**Special cases**

- If \( p \) is close to 1, then the second term is small, so we expect a *polynomial* time bound.
- If \( p \) is sufficiently small, then the second term is large, and we expect an *exponential* time bound.
- What happens for \( p \) in between?
AIM: A MORE PRECISE ANALYSIS OF THE EXHAUSTIVE ALGORITHM

MIS contains either \( v \) or not

\[
X_n \overset{d}{=} X_{n-1} + X^*_{n-1 - \text{Binom}(n-1; p)} \quad (n \geq 2),
\]

with \( X_0 = 0 \) and \( X_1 = 1 \).

Special cases

- If \( p \) is close to 1, then the second term is small, so we expect a polynomial time bound.
- If \( p \) is sufficiently small, then the second term is large, and we expect an exponential time bound.
- What happens for \( p \) in between?
AIM: A MORE PRECISE ANALYSIS OF THE EXHAUSTIVE ALGORITHM

MIS contains either \( v \) or not

\[
X_n \overset{d}{=} X_{n-1} + X^*_n - \text{Binom}(n-1; p) \quad (n \geq 2),
\]

with \( X_0 = 0 \) and \( X_1 = 1 \).

Special cases

- If \( p \) is close to 1, then the second term is small, so we expect a \textit{polynomial} time bound.
- If \( p \) is sufficiently small, then the second term is large, and we expect an \textit{exponential} time bound.
- What happens for \( p \) in between?
The expected value $\mu_n := \mathbb{E}(X_n)$ satisfies

$$\mu_n = \mu_{n-1} + \sum_{0 \leq j < n} \binom{n-1}{j} p^j q^{n-1-j} \mu_{n-1-j}. $$

with $\mu_0 = 0$ and $\mu_1 = 1$.

Poisson generating function

Let $\tilde{f}(z) := e^{-z} \sum_{n \geq 0} \frac{\mu_n z^n}{n!}$. Then

$$\tilde{f}'(z) = \tilde{f}(qz) + e^{-z}. $$
The expected value $\mu_n := \mathbb{E}(X_n)$ satisfies

$$\mu_n = \mu_{n-1} + \sum_{0 \leq j < n} \binom{n-1}{j} p^j q^{n-1-j} \mu_{n-1-j}.$$ 

with $\mu_0 = 0$ and $\mu_1 = 1$.

Poisson generating function

Let $\tilde{f}(z) := e^{-z} \sum_{n \geq 0} \mu_n z^n / n!$. Then

$$\tilde{f}'(z) = \tilde{f}(qz) + e^{-z}.$$
Laplace transform

The Laplace transform of $\tilde{f}$

$$\mathcal{L}(s) = \int_{0}^{\infty} e^{-sx}\tilde{f}(x) \, dx$$

satisfies

$$s\mathcal{L}(s) = \frac{1}{q} \mathcal{L}\left(\frac{s}{q}\right) + \frac{1}{s+1}.$$

Exact solutions

$$\mathcal{L}(s) = \sum_{j \geq 0} \frac{q^{j+1}}{s^{j+1}(s + q^j)}.$$
The Laplace transform of $\tilde{f}$

$$\mathcal{L}(s) = \int_0^\infty e^{-sx}\tilde{f}(x) \, dx$$

satisfies

$$s\mathcal{L}(s) = \frac{1}{q} \mathcal{L}\left(\frac{s}{q}\right) + \frac{1}{s+1}.$$
RESOLUTION OF THE RECURRENCE

Exact solutions

\[ \mathcal{L}(s) = \sum_{j \geq 0} \frac{q^{j+1}}{s^{j+1}(s + q^j)}. \]

Inverting gives

\[ \tilde{f}(z) = \sum_{j \geq 0} \frac{q^{j+1}}{j!} z^{j+1} \int_0^1 e^{-q^j uz} (1 - u)^j \, du. \]

Thus

\[ \mu_n = \sum_{1 \leq j \leq n} \binom{n}{j} (-1)^j \sum_{1 \leq \ell \leq j} (-1)^\ell q^{j(\ell-1)} \frac{1}{\binom{j+1}{2}}, \text{ or} \]

\[ \mu_n = n \sum_{0 \leq j < n} \binom{n-1}{j} q^{j+1} \sum_{0 \leq \ell < n-j} \binom{n-1-j}{\ell} q^{\ell} (1 - q^j)^{n-1-j-\ell} \frac{1}{j + \ell + 1}. \]

Neither is useful for numerical purposes for large \( n \).
RESOLUTION OF THE RECURRENCE

Exact solutions

\[ \mathcal{L}(s) = \sum_{j \geq 0} \frac{q^{j+1}}{s^{j+1}(s + q^j)}. \]

Inverting gives

\[ \tilde{f}(z) = \sum_{j \geq 0} \frac{q^{j+1}}{j!} z^{j+1} \int_{0}^{1} e^{-q^juz}(1 - u)^j \, du. \]

Thus

\[ \mu_n = \sum_{1 \leq j \leq n} \binom{n}{j} (-1)^j \sum_{1 \leq \ell \leq j} (-1)^\ell q^{j(\ell - 1) - \binom{\ell}{2}}, \text{ or} \]

\[ \mu_n = n \sum_{0 \leq j < n} \binom{n-1}{j} q^{j+1} \sum_{0 \leq \ell < n-j} \binom{n-1-j}{\ell} q^{\ell} (1 - q^j)^{n-1-j-\ell} \frac{q^{\ell}(1 - q^j)^{n-1-j-\ell}}{j + \ell + 1}. \]

Neither is useful for numerical purposes for large \( n \).
RESOLUTION OF THE RECURRENCE

Exact solutions

\[ \mathcal{L}(s) = \sum_{j \geq 0} \frac{q^{(j+1)/2}}{s^{j+1}(s + q^j)}. \]

Inverting gives

\[ \tilde{f}(z) = \sum_{j \geq 0} \frac{q^{(j+1)/2}}{j!} z^{j+1} \int_0^1 e^{-q^j uz} (1 - u)^j \, du. \]

Thus

\[ \mu_n = \sum_{1 \leq j \leq n} \binom{n}{j} (-1)^j \sum_{1 \leq \ell \leq j} (-1)^\ell q^{j(\ell-1) - \binom{\ell}{2}}, \text{ or} \]

\[ \mu_n = n \sum_{0 \leq j < n} \binom{n-1}{j} q^{(j+1)/2} \sum_{0 \leq \ell < n-j} \binom{n-1-j}{\ell} q^{\ell} (1 - q^j)^{n-1-j-\ell} \frac{q^{\ell}(1 - q^j)^{n-1-j-\ell}}{j + \ell + 1}. \]

Neither is useful for numerical purposes for large \( n \).
Back-of-the-envelope calculation

Take $q = 1/2$. Since Binom$(n - 1; \frac{1}{2})$ has mean $n/2$, we roughly have

$$\mu_n \approx \mu_{n-1} + \mu_{\lfloor n/2 \rfloor}.$$ 

This is reminiscent of Mahler’s partition problem. Indeed, if $\varphi(z) = \sum_n \mu_n z^n$, then

$$\varphi(z) \approx \frac{1 + z}{1 - z} \varphi(z^2) = \prod_{j \geq 0} \frac{1}{1 - z^{2^j}}.$$ 

So we expect that (de Bruijn, 1948; Dumas and Flajolet, 1996)

$$\log \mu_n \approx c \left( \log \frac{n}{\log_2 n} \right)^2 + c' \log n + c'' \log \log n + \text{Periodic}_n.$$
Take $q = 1/2$. Since $\text{Binom}(n - 1; \frac{1}{2})$ has mean $n/2$, we roughly have
\[
\mu_n \approx \mu_{n-1} + \mu_{\lfloor n/2 \rfloor}.
\]
This is reminiscent of Mahler’s partition problem. Indeed, if $\varphi(z) = \sum_n \mu_n z^n$, then
\[
\varphi(z) \approx \frac{1 + z}{1 - z} \varphi(z^2) = \prod_{j \geq 0} \frac{1}{1 - z^{2j}}.
\]
So we expect that (de Bruijn, 1948; Dumas and Flajolet, 1996)
\[
\log \mu_n \approx c \left( \log \frac{n}{\log_2 n} \right)^2 + c' \log n + c'' \log \log n + \text{Periodic}_n.
\]
Back-of-the-envelope calculation

Take $q = 1/2$. Since $\text{Binom}(n - 1; \frac{1}{2})$ has mean $n/2$, we roughly have

$$\mu_n \approx \mu_{n-1} + \mu_{\lfloor n/2 \rfloor}.$$ 

This is reminiscent of Mahler’s partition problem. Indeed, if $\varphi(z) = \sum_n \mu_n z^n$, then

$$\varphi(z) \approx \frac{1 + z}{1 - z} \varphi(z^2) = \prod_{j \geq 0} \frac{1}{1 - Z^{2^j}}.$$

So we expect that (de Bruijn, 1948; Dumas and Flajolet, 1996)

$$\log \mu_n \approx c \left( \log \frac{n}{\log_2 n} \right)^2 + c' \log n + c'' \log \log n + \text{Periodic}_n.$$
Poisson heuristic (de-Poissonization, saddle-point method)

\[
\mu_n = \frac{n!}{2\pi i} \int_{|z|=n} z^{-n-1} e^z \tilde{f}(z) \, dz \\
\approx \sum_{j \geq 0} \frac{\tilde{f}(j)(n)}{j!} \frac{n!}{2\pi i} \int_{|z|=n} z^{-n-1} e^z (z - n)^j \, dz \\
= \tilde{f}(n) + \sum_{j \geq 2} \frac{\tilde{f}(j)(n)}{j!} \tau_j(n),
\]

where \( \tau_j(n) := n! [z^n] e^z (z - n)^j = j! [z^j](1 + z)^n e^{-nz} \) (Charlier polynomials). In particular, \( \tau_0(n) = 1, \tau_1(n) = 0, \tau_2(n) = -n, \tau_3(n) = 2n, \) and \( \tau_4(n) = 3n^2 - 6n. \)
Poisson heuristic (de-Poissonization, saddle-point method)

\[ \mu_n = \frac{n!}{2\pi i} \oint_{|z|=n} z^{-n-1} e^z \tilde{f}(z) \, dz \]

\[ \approx \sum_{j \geq 0} \frac{\tilde{f}^{(j)}(n)}{j!} \frac{n!}{2\pi i} \oint_{|z|=n} z^{-n-1} e^z (z - n)^j \, dz \]

\[ = \tilde{f}(n) + \sum_{j \geq 2} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n), \]

where \( \tau_j(n) := n! [z^n] e^z (z - n)^j = j! [z^j] (1 + z)^n e^{-nz} \) (Charlier polynomials). In particular, \( \tau_0(n) = 1, \tau_1(n) = 0, \tau_2(n) = -n, \tau_3(n) = 2n, \) and \( \tau_4(n) = 3n^2 - 6n. \)
Asymptotics of $\tilde{f}(x)$

Let $\rho = 1 / \log(1/q)$ and $R \log R = x / \rho$. Then

$$
\tilde{f}(x) \sim \frac{R^{\rho+1/2} e^{(\rho/2)(\log R)^2} G(\rho \log R)}{\sqrt{2\pi \rho \log R}} \left( 1 + \sum_{j \geq 1} \frac{\phi_j(\rho \log R)}{(\rho \log R)^j} \right),
$$

as $x \to \infty$, where $G(u) := q^{(\{u\}^2 + \{u\})/2} F(q^{-\{u\}})$,

$$
F(s) = \sum_{-\infty < j < \infty} \frac{q^{j(j+1)/2}}{1 + q^j s} s^{j+1},
$$

and the $\phi_j(u)$’s are bounded, 1-periodic functions of $u$ involving the derivatives $F^{(j)}(q^{-\{u\}})$. 

Hsien-Kuei Hwang
A MORE EXPLICIT ASYMPTOTIC APPROXIMATION

\[ R = \frac{x}{\rho} / W\left(\frac{x}{\rho}\right), \text{Lambert’s } W\text{-function} \]

\[ W(x) = \log x - \log \log x + \frac{\log \log x}{\log x} + \frac{(\log \log x)^2 - 2 \log \log x}{2(\log x)^2} + \cdots. \]

So that

\[ \tilde{f}(x) \sim x^{\rho + 1/2} G\left(\rho \log \frac{x/\rho}{\log(x/\rho)}\right) \exp \left(\frac{\rho}{2} \left(\log \frac{x/\rho}{\log(x/\rho)}\right)^2\right). \]

Method of proof: a variant of the saddle-point method

\[ \tilde{f}(x) = \frac{1}{2\pi i} \int_{1-i0}^{1+i0} e^{sz} L(s) \, ds \]
A MORE EXPLICIT ASYMPTOTIC APPROXIMATION

\[ R = \frac{x}{\rho}/W(x/\rho), \text{ Lambert's } W\text{-function} \]

\[ W(x) = \log x - \log \log x + \frac{\log \log x}{\log x} + \frac{(\log \log x)^2 - 2 \log \log x}{2(\log x)^2} + \cdots. \]

So that

\[ \tilde{f}(x) \sim x^{\rho+1/2} G \left( \rho \log \frac{x/\rho}{\log(x/\rho)} \right) \exp \left( \frac{\rho}{2} \left( \log \frac{x/\rho}{\log(x/\rho)} \right)^2 \right). \]

Method of proof: a variant of the saddle-point method

\[ \tilde{f}(x) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} e^{sz} \mathcal{L}(s) \, ds \]
A MORE EXPLICIT ASYMPTOTIC APPROXIMATION

\[ R = \frac{x}{\rho} \big/ W\left( \frac{x}{\rho} \right), \] Lambert’s \( \mathcal{W} \)-function

\[ W(x) = \log x - \log \log x + \frac{\log \log x}{\log x} + \frac{(\log \log x)^2 - 2 \log \log x}{2(\log x)^2} + \cdots. \]

So that

\[ \tilde{f}(x) \sim \frac{x^{\rho + 1/2} G\left( \frac{x}{\rho} \log \frac{x}{\rho} \right)}{\sqrt{2\pi \rho^{\rho + 1/2} \log x}} \exp \left( \frac{\rho}{2} \left( \log \frac{x}{\rho} \log \frac{x}{\rho} \right)^2 \right). \]

Method of proof: a variant of the saddle-point method

\[ \tilde{f}(x) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} \mathcal{L}(s) \exp(sz) \, ds \]
Four properties are sufficient

The following four properties are enough to justify the Poisson-Charlier expansion.

- $\tilde{f}'(z) = \tilde{f}(qz) + e^{-z}$;
- $F(s) = sF(qs)$ ($F(s) = \sum_{i \in \mathbb{Z}} q^{i+1}/2 s^{i+1}/(1 + q^i s)$);
- $\tilde{f}(j)(x) \sim \left( \frac{\rho \log x}{x} \right)^j$;
- $|f(z)| \leq f(|z|)$ where $f(z) := e^z \tilde{f}(z)$.

Thus ($\rho = 1 / \log(1/q)$)

$$\mu_n \sim \frac{n^{\rho+1/2} G \left( \rho \log \frac{n/\rho}{\log(n/\rho)} \right)}{\sqrt{2\pi \rho^{\rho+1/2} \log n}} \exp \left( \frac{\rho}{2} \left( \log \frac{n/\rho}{\log(n/\rho)} \right)^2 \right).$$
Four properties are sufficient

The following four properties are enough to justify the Poisson-Charlier expansion.

- \( \tilde{f}'(z) = \tilde{f}(qz) + e^{-z} \);
- \( F(s) = sF(qs) \) \( F(s) = \sum_{i \in \mathbb{Z}} q^{(i+1)/2} s^{i+1} / (1 + q^i s) \);
- \( \tilde{f}^{(j)}(x) \sim \left( \frac{\rho \log x}{x} \right)^j \);
- \( |f(z)| \leq f(|z|) \) \text{ where } f(z) := e^z \tilde{f}(z) \).

Thus \( \rho = 1 / \log(1/q) \)

\[
\mu_n \sim \frac{n^{\rho+1/2} G \left( \rho \log \frac{n/\rho}{\log(n/\rho)} \right)}{\sqrt{2\pi \rho^{\rho+1/2} \log n}} \exp \left( \frac{\rho}{2} \left( \log \frac{n/\rho}{\log(n/\rho)} \right)^2 \right).
\]
\[ \sigma_n := \sqrt{\mathbb{V}(X_n)} \]

\[ \sigma_n^2 = \sigma_{n-1}^2 + \sum_{0 \leq j < n} \pi_{n,j} \sigma_{n-1-j}^2 + T_n, \quad \pi_{n,j} := \binom{n-1}{j} p^j q^{n-1-j}, \]

where \( T_n := \sum_{0 \leq j < n} \pi_{n,j} \Delta_{n,j}^2, \quad \Delta_{n,j} := \mu_j + \mu_{n-1} - \mu_n. \)

Asymptotic transfer: \( a_n = a_{n-1} + \sum_{0 \leq j < n} \pi_{n,j} a_{n-1-j} + b_n \)

If \( b_n \sim n^\beta (\log n)^\kappa \tilde{f}(n)^\alpha \), where \( \alpha > 1, \beta, \kappa \in \mathbb{R} \), then

\[ a_n \sim \sum_{j \leq n} b_j \sim \frac{n}{\alpha \rho \log n} b_n. \]
VARIANCE OF $X_n$

$$\sigma_n := \sqrt{\mathbb{V}(X_n)}$$

$$\sigma_n^2 = \sigma_{n-1}^2 + \sum_{0 \leq j < n} \pi_{n,j} \sigma_{n-1-j}^2 + T_n, \quad \pi_{n,j} := \binom{n-1}{j} p^j q^{n-1-j},$$

where $T_n := \sum_{0 \leq j < n} \pi_{n,j} \Delta_{n,j}^2$, $\Delta_{n,j} := \mu_j + \mu_{n-1} - \mu_n$.

Asymptotic transfer: $a_n = a_{n-1} + \sum_{0 \leq j < n} \pi_{n,j} a_{n-1-j} + b_n$

If $b_n \sim n^\beta (\log n)^\kappa \tilde{f}(n)^\alpha$, where $\alpha > 1$, $\beta, \kappa \in \mathbb{R}$, then

$$a_n \sim \sum_{j \leq n} b_j \sim \frac{n}{\alpha \rho \log n} b_n.$$
Asymptotics of $T_n$: by elementary means

$$T_n \sim q^{-1} p \rho^4 n^{-3} (\log n)^4 \tilde{f}(n)^2.$$ 

Applying the asymptotic transfer

$$\sigma_n^2 \sim C n^{-2} (\log n)^3 \tilde{f}(n)^2.$$ 

where $C := p \rho^3 / (2q)$. 

\[
\frac{\text{Variance}}{\text{Mean}^2} \sim C \frac{(\log n)^3}{n^2}
\]
Asymptotics of $T_n$: by elementary means

\[ T_n \sim q^{-1} p \rho^4 n^{-3} (\log n)^4 \tilde{f}(n)^2. \]

Applying the asymptotic transfer

\[ \sigma_n^2 \sim C n^{-2} (\log n)^3 \tilde{f}(n)^2. \]

where $C := p \rho^3 / (2q)$.

\[
\frac{\text{Variance}}{\text{Mean}^2} \sim C \frac{(\log n)^3}{n^2}
\]
Asymptotics of $T_n$: by elementary means

\[ T_n \sim q^{-1} p \rho^4 n^{-3} (\log n)^4 \tilde{f}(n)^2. \]

Applying the asymptotic transfer

\[ \sigma_n^2 \sim C n^{-2} (\log n)^3 \tilde{f}(n)^2. \]

where $C := \rho \rho^3 / (2q)$.

\[ \frac{\text{Variance}}{\text{Mean}^2} \sim C \frac{(\log n)^3}{n^2} \]
Convergence in distribution

The distribution of $X_n$ is asymptotically normal

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1),$$

with convergence of all moments.

Proof by the method of moments

- Derive recurrence for $\mathbb{E}(X_n - \mu_n)^m$.
- Prove by induction (using the asymptotic transfer) that

$$\mathbb{E}(X_n - \mu_n)^m \begin{cases} \sim \frac{(m)!}{(m/2)!2^{m/2}} \sigma_n^m, & \text{if } 2 \mid m, \\ = o(\sigma_n^m), & \text{if } 2 \nmid m, \end{cases}$$
Convergence in distribution

The distribution of $X_n$ is asymptotically normal

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1),$$

with convergence of all moments.

Proof by the method of moments

- Derive recurrence for $\mathbb{E}(X_n - \mu_n)^m$.
- Prove by induction (using the asymptotic transfer) that

$$\mathbb{E}(X_n - \mu_n)^m \begin{cases} 
\sim \frac{(m)!}{(m/2)!2^{m/2}} \sigma_n^m, & \text{if } 2 \mid m, \\
= o(\sigma_n^m), & \text{if } 2 \nmid m,
\end{cases}$$
A STRAIGHTFORWARD EXTENSION

\[ b = 1, 2, \ldots \]

\[ X_n \overset{d}{=} X_{n-b} + X^n_{n-b} \text{Binom}(n-b;p), \]

with \( X_n = 0 \) for \( n < b \) and \( X_b = 1 \).

For example, MAXIMUM TRIANGLE PARTITION:

\[ X_n \overset{d}{=} X_{n-3} + X^n_{n-3} \text{Binom}(n-3;p^3), \]

The same tools we developed apply

\( X_n \) asymptotically normally distributed with mean and variance of the same order as the case \( b = 1 \).
A STRAIGHTFORWARD EXTENSION

\[ b = 1, 2, \ldots \]

\[ X_n \overset{d}{=} X_{n-b} + X^*_{n-b} - \text{Binom}(n-b; p), \]

with \( X_n = 0 \) for \( n < b \) and \( X_b = 1 \).

For example, MAXIMUM TRIANGLE PARTITION:

\[ X_n \overset{d}{=} X_{n-3} + X^*_{n-3} - \text{Binom}(n-3; p^3), \]

The same tools we developed apply

\( X_n \) asymptotically normally distributed with mean and variance of the same order as the case \( b = 1 \).
A STRAIGHTFORWARD EXTENSION

\[ b = 1, 2, \ldots \]

\[ X_n \overset{d}{=} X_{n-b} + X^*_{n-b} \text{ Binom}(n-b;p), \]

with \( X_n = 0 \) for \( n < b \) and \( X_b = 1 \).

For example, MAXIMUM TRIANGLE PARTITION:

\[ X_n \overset{d}{=} X_{n-3} + X^*_{n-3} \text{ Binom}(n-3;p^3), \]

The same tools we developed apply

\( X_n \) asymptotically normally distributed with mean and variance of the same order as the case \( b = 1 \).
What happens if $X_n \overset{d}{=} X_{n-1} + X_{\text{uniform}[0,n-1]}$?

\[ \mu_n = \mu_{n-1} + \frac{1}{n} \sum_{0 \leq j < n} \mu_j, \]

satisfies $\mu_n \sim c n^{-1/4} e^{2\sqrt{n}}$. Note: $\mu_n \approx \mu_{n-1} + \mu_{n/2}$ fails.

Limit law not Gaussian (by method of moments)

\[ \frac{X_n}{\mu_n} \overset{d}{\to} X, \]

where $g(z) := \sum_{m \geq 1} \mathbb{E}(X^m) z^m / (m \cdot m!)$ satisfies

\[ z^2 g''' + zg' - g = zg'. \]
A NATURAL VARIANT

What happens if \( X_n \overset{d}{=} X_{n-1} + X^\ast_{\text{uniform}[0,n-1]} \)?

\[
\mu_n = \mu_{n-1} + \frac{1}{n} \sum_{0 \leq j < n} \mu_j,
\]

satisfies \( \mu_n \sim cn^{-1/4}e^{2\sqrt{n}} \). Note: \( \mu_n \approx \mu_{n-1} + \mu_{n/2} \) fails.

Limit law not Gaussian (by method of moments)

\[
\frac{X_n}{\mu_n} \overset{d}{\to} X,
\]

where \( g(z) := \sum_{m \geq 1} \mathbb{E}(X^m)z^m/(m \cdot m!) \) satisfies

\[
z^2g'' + zg' - g = zgg'.
\]
What happens if $X_n \overset{d}{=} X_{n-1} + X_{\text{uniform}[0,n-1]}$?

$$\mu_n = \mu_{n-1} + \frac{1}{n} \sum_{0 \leq j < n} \mu_j,$$

satisfies $\mu_n \sim cn^{-1/4} e^{2\sqrt{n}}$. **Note:** $\mu_n \approx \mu_{n-1} + \mu_{n/2}$ fails.

Limit law not Gaussian (by method of moments)

$$\frac{X_n}{\mu_n} \overset{d}{\rightarrow} X,$$

where $g(z) := \sum_{m \geq 1} \mathbb{E}(X^m) z^m / (m \cdot m!) \text{ satisfies } z^2 g'' + zg' - g = zgg'$. 

Hsien-Kuei Hwang

PROBABILISTIC ANALYSIS OF AN EXHAUSTIVE SEARCH ALGORITHM IN RANDOM GRAPHS
Random graph algorithms: a rich source of interesting recurrences

Obrigado!
Random graph algorithms: a rich source of interesting recurrences

Obrigado!