PROBABILISTIC ANALYSIS OF AN EXHAUSTIVE SEARCH ALGORITHM IN RANDOM GRAPHS

Vytas Zacharovas

Academia Sinica, Taiwan
(joint work with Hsien-Kuei Hwang, Cyril Banderier and Vlady Ravelomanana)

Workshop on Discrete Mathematics, Vienna, Austria
November 19-22, 2008
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.

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

**NP hard!!**
MAXIMUM INDEPENDENT SET

Equivalent version
The same problem as **MAXIMUM CLIQUE** on the complementary graph (clique = complete subgraph).
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), Bollobás 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 $q = 1 - p$. 
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 $(q := 1 - p)$

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

where $q = 1 - p$. 
Problem
Find such an independent set $A$ in $G$ that no other node from $G$ can be added to $A$ without destroying the independence of $A$.

Solution
Initially $A = \emptyset$.

- Chose $v \in G$.
- $A := A \cup \{v\}$, $G := G \ast v$, where $G \ast v$ is the graph obtained from $G$ by deleting node $v$ together with all its neighboring nodes and their edges.
- Continue until $G = \emptyset$. 
DESCRIPTION OF THE GREEDY ALGORITHM

Problem
Find such an independent set \( A \) in \( G \) that no other node from \( G \) can be added to \( A \) without destroying the independence of \( A \).

Solution
Initially \( A = \emptyset \).

- Chose \( v \in G \).
- \( A := A \cup \{v\} \), \( G := G \ast v \), where \( G \ast v \) is the graph obtained from \( G \) by deleting node \( v \) together with all its neighboring nodes and their edges.
- Continue until \( G = \emptyset \).
DESCRIPTION OF THE GREEDY ALGORITHM

Problem
Find such an independent set $A$ in $G$ that no other node from $G$ can be added to $A$ without destroying the independence of $A$.

Solution
Initially $A = \emptyset$.

- Chose $v \in G$.
- $A := A \cup \{v\}$, $G := G \ast v$, where $G \ast v$ is the graph obtained from $G$ by deleting node $v$ together with all its neighboring nodes and their edges.
- Continue until $G = \emptyset$. 
DESCRIPTION OF THE GREEDY ALGORITHM

Problem
Find such an independent set $A$ in $G$ that no other node from $G$ can be added to $A$ without destroying the independence of $A$.

Solution
Initially $A = \emptyset$.

- Chose $v \in G$.
- $A := A \cup \{v\}$, $G := G * v$, where $G * v$ is the graph obtained from $G$ by deleting node $v$ together with all its neighboring nodes and their edges.
- Continue until $G = \emptyset$. 
DESCRIPTION OF THE GREEDY ALGORITHM

Problem
Find such an independent set $A$ in $G$ that no other node from $G$ can be added to $A$ without destroying the independence of $A$.

Solution
Initially $A = \emptyset$.

▶ Chose $v \in G$.

▶ $A := A \cup \{v\}$, $G := G \ast v$, where $G \ast v$ is the graph obtained from $G$ by deleting node $v$ together with all its neighboring nodes and their edges.

▶ Continue until $G = \emptyset$. 
A GREEDY ALGORITHM

Recurrence
The size of the resulting independent set $S_n$ satisfies recurrence relation:

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

with $S_0 \equiv 0$. 
ANALYSIS OF THE GREEDY ALGORITHM

Relatively easy

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

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

ANALYSIS OF THE GREEDY ALGORITHM

Relatively easy

- **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:**

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

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

▶ **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 $\left(\frac{1}{2} + \varepsilon\right)|\text{MIS}_n|$ 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 \((1/2 + \varepsilon)|MIS_n|\) whp or show that such an algorithm does not exist modulo some reasonable conjecture in the theory of computational complexity such as, e.g., \(P \neq NP\).
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 \((\frac{1}{2} + \varepsilon) |MIS_n|\) whp or show that such an algorithm does not exist modulo some reasonable conjecture in the theory of computational complexity such as, e.g., \(P \neq NP\).
POSITIVE RESULTS

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

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

– Pittel (1982):

\[
P\left(n^{\frac{1}{2} - \log_{1/2} n} \leq \text{Time used by Chvátal’s algo} \leq n^{\frac{1}{2} + \log_{1/2} n}\right) \geq 1 - e^{-c\log^2 n}
\]
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

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

- Pittel (1982):

$$P\left(n^{\frac{1-\varepsilon}{4} \log_{1/q} n} \leq \text{Time used by Chvátal’s algo} \leq n^{\frac{1+\varepsilon}{2} \log_{1/q} n}\right) \geq 1 - e^{-c \log^2 n}$$
POSITIVE RESULTS

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

- Chvátal (1977) proposes exhaustive algorithms where almost all $G_{n,1/2}$ creates at most $n^{2(1+\log_2 n)}$ subproblems.
- Pittel (1982):

$$\Pr \left( n^{\frac{1-\epsilon}{4} \log_{1/q} n} \leq \text{Time used by Chvátal’s algo} \leq n^{\frac{1+\epsilon}{2} \log_{1/q} n} \right) \geq 1 - e^{-c \log^2 n}$$
Problem

Suppose we want to compute the stability number of the graph $G$, that is $|MIS(G)|$.

- Choose a node $v \in G$.
- Delete from $G$ the node $v$ together with all its edges, that is obtain graph $G - v$. Compute $|MIS(G - v)|$.
- Delete from $G$ the node $v$ together with all its neighboring nodes and their edges. The obtained graph denote by $G \ast v$. Compute $|MIS(G \ast v)|$.

$|MIS(G)| = \max\{|MIS(G - v)|, |MIS(G \ast v)| + 1\}$. 
PROBLEM

Suppose we want to compute the stability number of the graph $G$, that is $|MIS(G)|$.

1. **Choose a node** $v \in G$.
2. **Delete from $G$ the node $v$ together with all its edges**, that is obtain graph $G - v$. **Compute** $|MIS(G - v)|$.
3. **Delete from $G$ the node $v$ together with all its neighboring nodes and their edges**. The obtained graph denote by $G \ast v$. **Compute** $|MIS(G \ast v)|$.
4. $|MIS(G)| = \max\{|MIS(G - v)|, |MIS(G \ast v)| + 1\}$.
DESCRIPTION OF THE EXHAUSTIVE ALGORITHM PROPOSED BY V. CHVATAL

Problem
Suppose we want to compute the stability number of the graph $G$, that is $|\text{MIS}(G)|$.

- Chose a node $v \in G$.
- Delete from $G$ the node $v$ together with all its edges, that is obtain graph $G - v$. Compute $|\text{MIS}(G - v)|$.
- Delete from $G$ the node $v$ together with all its neighboring nodes and their edges. The obtained graph denote by $G \ast v$. Compute $|\text{MIS}(G \ast v)|$.
- $|\text{MIS}(G)| = \max\{|\text{MIS}(G - v)|, |\text{MIS}(G \ast v)| + 1\}$. 

DESCRIPTION OF THE EXHAUSTIVE ALGORITHM PROPOSED BY V. CHVATAL

Problem
Suppose we want to compute the stability number of the graph $G$, that is $|MIS(G)|$.

- Chose a node $v \in G$.
- Delete from $G$ the node $v$ together with all its edges, that is obtain graph $G - v$. Compute $|MIS(G - v)|$.
- Delete from $G$ the node $v$ together with all its neighboring nodes and their edges. The obtained graph denote by $G \ast v$. Compute $|MIS(G \ast v)|$.
- $|MIS(G)| = \max\{ |MIS(G - v)|, |MIS(G \ast v)| + 1 \}$. 
AIM: A MORE PRECISE ANALYSIS OF THE EXHAUSTIVE ALGORITHM

The time needed to complete the algorithm $X_n$ is a random variable satisfying recurrence relation

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

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?
The time needed to complete the algorithm $X_n$ is a random variable satisfying recurrence relation

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

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

The time needed to complete the algorithm $X_n$ is a random variable satisfying recurrence relation

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

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

The time needed to complete the algorithm $X_n$ is a random variable satisfying recurrence relation

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

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?
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

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

$$
\tilde{f}'(z) = \tilde{f}(qz) + e^{-z}.
$$
RESOLUTION OF THE RECURRENCE

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

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

The Laplace transform of $\tilde{f}$

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

satisfies

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

Exact solutions

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

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

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

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

Exact solutions

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

Inverting gives

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

Thus

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

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

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

Exact solutions

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

Inverting gives \( \tilde{f}(z) = \sum_{j \geq 0} \frac{q^{\left(\frac{j+1}{2}\right)}}{j!} z^{j+1} \int_{0}^{1} e^{-q^j uz}(1 - u)^j \, du. \)

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

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

Neither is useful for numerical purposes for large \( n. \)
ASYMPTOTICS OF $\mu_n$

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 $\mu_n$

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 $\tau_4(n) = 3n^2 - 6n$. 
A MORE PRECISE EXPANSION FOR $\tilde{f}(x)$

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 $F^{(j)}(q^{-\{u\}}).$
A MORE EXPLICIT ASYMPTOTIC APPROXIMATION

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

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

So that

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

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

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

\[ R = x/\rho / W(x/\rho) \], Lambert’s \( W \)-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

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

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

\[ \tilde{f}(x) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} e^{xs} \mathcal{L}(s) \, ds \]
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

\[ \tilde{f}(x) = \frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} e^{xs} \mathcal{L}(s) \, ds \]
JUSTIFICATION OF THE POISSON HEURISTIC

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

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

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

\[
\mu_n \sim \frac{n^{\rho+1/2} G \left( \rho \log \frac{n/\rho}{\log(n/\rho)} \right)}{\sqrt{2\pi\rho^{\rho+1/2} \log n}} \exp \left( \frac{\rho}{2} \left( \log \frac{n/\rho}{\log(n/\rho)} \right)^2 \right).
\]
JUSTIFICATION OF THE POISSON HEURISTIC

Four properties are sufficient

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

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

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

\[
\mu_n \sim \frac{n^{\rho+1/2} G\left( \rho \log \frac{n/\rho}{\log(n/\rho)} \right)}{\sqrt{2\pi \rho^{\rho+1/2} \log n}} \exp \left( \frac{\rho}{2} \left( \log \frac{n/\rho}{\log(n/\rho)} \right)^2 \right).
\]
Rough estimates

Corollary

Thus we have

$$\mathbb{E}X_n \asymp n^{\rho + 1/2} \exp \left( \frac{\rho}{2} \left( \log \frac{n/\rho}{\log(n/\rho)} \right)^2 \right),$$

where

$$\rho = \rho(p) = \frac{1}{\log \frac{1}{1-p}}.$$ 

Compare with the result of Pittel (1982)

$$\mathbb{P} \left( n^{\frac{1-\varepsilon}{4}} \log_{1/q} n \leq \text{Time used by Chvátal's algo} \leq n^{\frac{1+\varepsilon}{2}} \log_{1/q} n \right) \geq 1 - e^{-c \log^2 n}$$
If we take $n = 300$ then according to our result for $p = 0.4$

$$\mathbb{E}X_n \approx 1.12 \cdot 10^{11}$$

while for $p = 0.6$

$$\mathbb{E}X_n \approx 3.38 \cdot 10^{7}$$

This means that our algorithm for $p = 0.6$ runs almost 3300 times faster than for $p = 0.4$. 

n=300
Idealized model

Dependence of $X_n$

Unfortunately $X_n$ in the recurrence

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

with $X_0 = 0$ and $X_1 = 1$, are not independent!

Idealized model

What will happen if we assume that $X_n$ are independent?
VARIANCE OF $X_n$ under the assumption of independence

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

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

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

$$a_n \sim \sum_{j \leq n} b_j \sim \frac{n}{\alpha \rho \log n} b_n.$$
VARIANCE OF $X_n$ under the assumption of independence

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

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

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 THE VARIANCE

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

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

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

$$\frac{\text{Variance}}{\text{Mean}^2} \sim C \frac{(\log n)^3}{n^2}$$
ASYMPTOTIC NORMALITY OF $X_n$

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

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

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

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

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

with \( X_n = 0 \) for \( n < b \) and \( X_b = 1 \).
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 $\mu_n \sim cn^{-1/4} e^{2\sqrt{n}}$.

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

$$
z^2 g'' + zg' - g = zgg'.
$$
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 $\mu_n \sim cn^{-1/4}e^{2\sqrt{n}}$.

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

$$z^2 g'' + zg' - g = zgg'.$$
Random graph algorithms: a rich source of interesting recurrences