# The differential equation method

Laura Eslava. IIMAS-UNAM
BUC-Chile

## Describing elements of a random graph

Consider a graph $G$ on $n$ vertices where edges are present independently at random

Vertex set $[n]=\{1,2,\dots,n\}$

Edges r.v.'s $X_{ij}=\mathbb{1}_{\{i\leftrightarrow j\}}$ independent

Some interesting statistics:

- number of edges
- isolated vertices
- number of components (of given size)
- degree sequence
- chromatic number
- independence number

Intuitively changing one variable in $\{X_{ij}\}$ won't change statistics by much.

This course: conditions to 'see' trends ... in a (graph) discrete process

## Chernoff bounds $t>0$

$$\mathbb{P}(X\geq a)=\mathbb{P}(e^{tX}\geq e^{ta})\leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

$$t=4a/\Sigma(b_i-a_i)^2 \qquad \text{optimize } t$$

If $S_n=\sum_{i=1}^n X_i$ independent $X_i$

$$\mathbb{P}(S_n\geq \mathbb{E}[S_n]+\varepsilon)\leq e^{-t\varepsilon}\prod_{i=1}^n \mathbb{E}[e^{t(X-\mathbb{E}[X])}]$$

+ bounded $X_i$ so that $a_i\leq X_i\leq b_i$

$$\mathbb{P}(S_n\geq \mathbb{E}[S_n]+\varepsilon)\leq e^{-2\varepsilon^2/\Sigma(b_i-a_i)^2}$$

This gives quantitative bounds for LLN $\mathbb{P}(X_i\in[0,1])$

$$\sum_{n\geq 1}\mathbb{P}\left(\left|\frac{S_n-\mathbb{E}[S_n]}{n}\right|>u\right)\leq \sum_{n\geq 1} 2e^{-nu^2/2}<\infty$$

$$\underset{\text{B.C.}}{\Longrightarrow} \mathbb{P}\left(\left|\frac{S_n-\mathbb{E}[S_n]}{n}\right|>u \text{ infinitely often}\right)=0$$

$$\frac{S_n-\mathbb{E}[S_n]}{n}\xrightarrow{a.s.}0$$

# Two generalization

① If $\{Z_i\}_{i\geq 0}$ martingale, $Z_0 = 0$

$|Z_{i+1} - Z_i| < c_i$

$$\mathbb{P}(Z_n \geq \mathbb{E}[Z_n] + \varepsilon) \leq e^{-2\varepsilon^2/\sum c_i^2}$$

Proof Sketch: $\mathcal{F}_i = \sigma\text{-algebra}(Z_1 \ldots Z_i)$

$$Z_n = \sum_{i=1}^{n} Z_i - Z_{i-1} = \sum_{i=1}^{n} V_i$$

$$\mathbb{E}[e^{tZ_n}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tV_i}\right]$$

$$= \mathbb{E}\left[\prod_{n-1}^{n-1} e^{tV_i}\, \mathbb{E}[e^{tV_n} \mid \mathcal{F}_{n-1}]\right]$$

Now $V_n$ satisfies conditions of Hoeffding's lemma.

② Let $\{X_i\}$ independent $X_i \in \mathcal{X}$

$$f : \mathcal{X}^n \to \mathbb{R}$$

Conditions

$$|f(x_1, x_2, \ldots x_n) - f(y_1, \ldots, y_n)| \leq \sum c_i \mathbb{1}_{\{x_i \neq y_i\}}$$

$$\mathbb{P}(f(X) \geq \mathbb{E}[f] + \varepsilon) \leq e^{-2\varepsilon^2/\sum c_i^2}$$

$$V_i = \mathbb{E}[f \mid \mathcal{F}_i] - \mathbb{E}[f \mid \mathcal{F}_{i-1}]$$

$$\mathcal{F}_i = \sigma\text{-alge}(X_1 \ldots X_i)$$

$\exists$ r.v.'s $L_i, U_i$ measurable wrt $\mathcal{F}_{i-1}$

$$L_i \leq V_i \leq U_i \qquad U_i - L_i \leq c_i$$

so can apply Hoeffding's lemma

Extension in Hamming distance

$$\underline{X} = (X_1, \dots, X_n) \in S^n \qquad A \subset S^n$$

Hamming dist $d_H(\underline{x}, \underline{y}) = \sum \mathbb{1}_{\{x_i \neq y_i\}}$

$$d_H(\underline{x}, A) = \inf_{y \in A} d_H(\underline{x}, \underline{y})$$

$$[A]_r = \{ \underline{x} \in S^n : d_H(\underline{x}, A) \leq r \}$$

How big becomes $[A]_{\varepsilon n}$

$$\mathbb{P}(\underline{X} \in [A]_{\varepsilon n}) = \mathbb{P}(d(\underline{X}, A) \leq \varepsilon n)$$
$$= 1 - \mathbb{P}(d(\underline{X}, A) > \varepsilon n)$$

* $\varepsilon n > \mathbb{E}[d(\underline{X}, A)]$ $\qquad u = \varepsilon n - \mathbb{E}[d(X, A)]$

$$\mathbb{P}(d(\underline{X}, A) > \varepsilon n) = \mathbb{P}(d(\underline{X}, A) > \mathbb{E}[\cdot] + u) \leq e^{-2nu^2}$$

Problem: $\mathbb{E}[d(\underline{X}, A)]$ unknown

$$\varepsilon n \geq B \geq \mathbb{E}[d(\underline{X}, A)]$$

$$\mathbb{P}(d(\underline{X}, A) \geq \varepsilon n) \leq \mathbb{P}(d(\underline{X}, A) \geq \mathbb{E}[\cdot] + u)$$

$$u = \varepsilon n - B$$

Trick: from Hoeffding lemma' proof

$$\mathbb{E}\left[e^{t(f - \mathbb{E}[f])}\right] \leq e^{nt^2/8}$$

$$f = -d(\underline{X}, A) \qquad \mathbb{E}\left[e^{-t\mathbb{E}[f]}\right] = e^{-t\mathbb{E}[f]}$$

$$\mathbb{P}(A) = \mathbb{E}\left[\mathbb{1}_{\{A\}} e^{tf}\right] \leq \mathbb{E}\left[e^{tf}\right] = \mathbb{E}\left[e^{-t \cdot d(X, A)}\right]$$

$$\mathbb{P}(A)\, e^{-t\mathbb{E}d(X,A)} \leq e^{nt^2/8}$$

Take logarithm + optimize $t$

$$\mathbb{E}[d(\underline{X}, A)] \leq \sqrt{\frac{-n \log \mathbb{P}(A)}{2}} = B$$

# Examples of analysis of algorithms

## ① Matchings in $G_{n,c/n}$ (Erdös-Rényi)

Graph on $n$ vertices edges independently present w.p. $\frac{c}{n}$

**Matching**: set of edges that do not share common vertices

**Algorithm:**

1. Delete isolated vertices
2. Select a uniformly random vertex of min-degree $v$
3. Include in matching one of its edges
4. Delete $v$ and all other incident edges.

**Thm**: $c>0$. The algorithm obtains a maximum matching a.a.s.

$(\mathbb{P}(\leftrightarrow) \to 1 \quad n \to \infty)$

## ② 3-coloring regular $G_{n,d}$

uniformly chosen graph on those with n vertices and $\deg(v)=d$ all v.

Proper coloring: assign colors to vertices so that no two adjacent vertices have same color

Algorithm: Classify uncolored vertices according to #available color

$S_i(t) = \{v : i$ possible colors for $v\}$

while $S_0(t)=0$

1. Select minimum $i$ with $S_i(t)>0$
2. select uniform vertex in $S_i(t)$
3. Color v randomly, update $S_i(t+1)$.

Failure if $S_0(t)>0$

Thm: $d < 4.003$. The algorithm obtains a proper 3-col. a.a.s.

## ③ 3-SAT formulae

$$(\bar{x}_1 \vee x_1 \vee x_2) \wedge \underbrace{(\bar{x}_2 \vee \bar{x}_1 \vee x_3)}_{\text{clause}}$$

literal: + or − variable

$$0 \vee 1 = 1$$

$$0 \wedge 1 = 0$$

Assign values to variables, to satisfy formula

Pure literal: when formula only contains literals $x_i$ (or only $\bar{x}_i$)

Algorithm: while $\exists$ pure literal available

1. choose one at random $x_i$ (or $\bar{x}_i$)
2. assign 'positive' value $x_i = 1$ (or $x_i = 0$)
3. remove all clauses containing $x_i$

## ③ 3-SAT formulae

**Pure literal:** when formula only contains literals $x_i$ (or only $\bar{x}_i$)

**Algorithm:** while $\exists$ pure literal available

1. choose one at random $x_i$ (or $\bar{x}_i$)
2. assign 'positive' value $x_i = 1$ (or $x_i = 0$)
3. remove all clauses containing $x_i$

3 literals per clause

$m$ clause

$n$ variables

**density** $\frac{m}{n}$

**Thm:** $\exists\, d_c > 0$. Construct a random 3-SAT where (# literals $= x_i$) $\simeq \mathrm{Poi}(\lambda_i)$ and density is $d$:

$d < d_c$ formula is satisf. a.a.s.

$d > d_c$ formula is not satisf. a.a.s.

# An introductory example

There are $n$ bins and $m = cn$ balls, sequentially place balls independ. into bins

How many $\underbrace{\text{empty bins}}_{B}$ there are left?

$$B = \sum_{j=1}^{n} B_j \qquad B_j = \mathbb{1}_{\{j\text{-th bin is empty}\}} \qquad \mathbb{P}(B_j = 1) = (1 - 1/n)^{cn} \Rightarrow \mathbb{E}[B] = n e^{-c}(1+o(1))$$

Let $\{X_i\}$ iid. $X_i \sim \mathrm{Unif}[n] \longrightarrow X_i = j$ if $i$-th ball goes to $j$-th bin

Now $B$ satisfies the bounded differences condition

$$\mathbb{P}(|B - \mathbb{E}[B]| \geq \varepsilon n) \leq 2 e^{-2\varepsilon^2 n^2 / \sum_{i=1}^{cn} 1^2} = 2 e^{-2n\varepsilon^2/c}$$

If need to estimate $\mathbb{E}[B]$

$Y_i$ = # empty bins when $i$ balls have been located

$$|Y_{i+1} - Y_i| \leq 1$$

If need to estimate $\mathbb{E}[B]$

$Y_i$ = # empty bins when $i$ balls have been located

Local changes, given current information $\mathcal{F}_i$ = generated by "info" at time $i$

$$\mathbb{E}[Y_{i+1} - Y_i \mid \mathcal{F}_i] = -1\,\mathbb{P}\left(\begin{array}{c}\text{ball into}\\ \text{empty bin}\end{array}\right) = \frac{-Y_i}{n}$$

Idealized ODE:

$$\frac{Y(tn)}{n} \approx y(t) \qquad \frac{Y(0)}{n} = y(0) = 1 \qquad \Delta Y_i = Y_{i+1} - Y_i$$

$$\frac{\mathbb{E}\left[\frac{\Delta Y(tn)}{n} \mid \mathcal{F}_{tn}\right]}{1/n} \cong y'(t) = -y(t)$$

Solution $y(t) = e^{-t}$

$$\frac{\mathbb{E}[B]}{n} = \mathbb{E}\left[\frac{Y_{cn}}{n}\right] \approx e^{-c}$$

# A toy deterministic question

How much can two collections of functions can differ if they have similar derivatives and initial values?

$\exists\, \lambda, \delta > 0$ small perturbations such that

$y_k(0) = \hat{y}_k$ cte $\qquad\qquad y_k'(t) = F_k(t, \underline{y(t)})$

$|z_k(0) - y_k(0)| \leq \lambda \qquad\qquad |z_k'(t) - F_k(t, \underline{z(t)})| \leq \delta$

Condition: $F$ is $L$-Lipschitz cont. in norm $\|\cdot\|_\infty$

in particular $\quad |F_k(t, y(t)) - F_k(t, z(t))| \leq L \max\limits_{k \in I} |y_k(t) - z_k(t)|$

$$\Longrightarrow \max_{k \in I} |y_k(t) - z_k(t)| \leq (\lambda + \delta T)\, e^{Lt} \qquad \forall t \leq T$$

$$|y'(s) - z'(s)| < |y'(s) - F(s, z(s))| + |\underbrace{F(s,z(s)) - z'(s)}| $$

$$< |F(s,y(s)) - F(s,z(s))| + \delta$$

$$\leq L \cdot |y(s) - z(s)| + \delta$$

$$\underbrace{|y(t) - z(t)|}_{x(t)} \leq |y(0) - z(0)| + \int_0^t |y'(s) - z'(s)| ds$$

$$\leq (\lambda + \delta T) + \int_0^t L\, x(s)\, ds$$

By Grönwall's ineq. $\Rightarrow$ $x(t) \leq (\lambda + \delta T)\, e^{Lt}$