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 $t > 0$

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

$$t = 4a / \Sigma(b_i - a_i)^2$$

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 $\quad \mathbb{P}(X_i \in [0,1])$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$u = \varepsilon n - B$$

Trick: from Hoeffding lemma' proof

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

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

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

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

Take logarithm + optimize t

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