理论分析意义

随机算法，不能靠跑1w次求平均

让用户放心

理论指导设计更好算法

Schema theorem

> 局限性：只关注了算法单步的变化

schema：解空间H

		$o(H)$	$d(H)$
e.g.	01#1#	3	3
	#1#1#	2	2
	###1#	1	0

- #：any
- order：确定位
- defining length：确定位距离的最大值

建模为马尔可夫链

链状态对应了算法的种群

convergence收敛性

无穷时间能否收敛

bit-wise mutation就是全局

时间复杂度

用户更关注，理论分析最关注

理论分析背景知识

全概率公式推下界

伪不等式

马尔可夫

[Markov's inequality] Let X be a random variable taking only non-negative values, and $E[X]$ its expectation. For any $t > 0$,

$$P(X \ge t) \le E[X]/t$$

Chernoff

> 用马尔可夫证明
>
> 所以任何用chernoff，都可以用马尔可夫

- 泊松分布

[Chernoff bounds] Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials, and $X = \sum_{i=1}^{n} X_i$. For any $\delta > 0$,

$$P(X \ge (1+\delta)E[X]) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[X]}$$

$$P(X \le (1-\delta)E[X]) \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{E[X]}$$

例子

：01采样，1个数不超过2n/3的概率

$$p(X \ge \tfrac{2}{3}n) \le \frac{\frac{n}{2}}{\frac{2}{3}n}$$

答案1：>=1/4

$$p(X \ge \tfrac{2}{3}n) = p(X \ge (1+\tfrac{1}{3})\tfrac{n}{2}) \le \left(\frac{e^{\frac{1}{3}}}{(\frac{4}{3})^{\frac{4}{3}}}\right)^{\frac{n}{2}} = c^{-\frac{n}{2}}$$

答案2：>=1-c^{n/2}

> chernoff得到了更紧的界，是因为马尔可夫用的不够巧妙

Union bound

01串通过bit-wise mutation把0个数减少j的概率

$$\le P\left(\bigcup_{i\ge 1} E_i\right) \le \binom{i}{j}\boxed{\left(\frac{1}{n}\right)^{j}} \rightarrow P(E_i)$$