随机算法复习纲要

1. 绪论

- Max-3-CNF 问题
- 动机：
  - 日常性：微信红包（金额服从正态分布）
  - 简洁高效性：文档去重（k-gram 数字指纹）
  - 必须性：大数据可视化（抽样算法）
  - 创新必备性：PageRank（平稳分布、分布式计算）
  - 广泛性：数据降维、集合配置（完全随机生成，这也行，纳尼？）、物理化学实验
- 集合相似性连接
  - 相似度
  - minHash
  - LSH

2. 随机算法及其分类

- 概率基础
  - 概率空间、容斥原理、合并界限、条件概率条件概率链、全概率公式、随机变量、独立性、数学期望、期望的线性性质、方差、二项分布
  - Markov 不等式（积分证明）
  - Chebyshev 不等式（利用 Markov 不等式证明）
  - 尾概率界（Tail Bound）
- 数值随机算法
  - 计算Π值
  - 计算定积分
- 随机选择算法
  - LAZYSELECT(S, k)算法及其性能分析
  - Las Vegas 算法
- 素数测试随机算法
  - 算法设计（N 进行 m 次测试）、性能分析、蒙特卡罗算法
  - 概率放大
  - Monte Carlo vs. Las Vegas
- 随机排序算法
  - 划分、递归求解算法及其复杂度分析（O(n log n)）
  - 舍伍德算法
- 最小割随机算法
  - CONTRACTION 算法及其复杂度分析（O(n^2)，正确率 2/n^2 较低，错

误率 1-2/n^2 较高）
- 概率放大后的算法 Amplify 及其复杂度分析（O(n^4)，错误率 1/e）
- 混合精确算法得到 DetRan、Amplify2 及其复杂复分析（O(n^8/3)，错误率 1/e）
- 递归算法 RepTree 及其复杂度分析（O(n^2 log^2 n)，1/e）

3. 随机算法复杂度下界

- NOR 电路估值问题
  - 确定性算法（O(2^k)）
  - 深度优先估值、随机估值算法 RandEval 及其时间复杂度引理（O(n^0.694)）
  - 博弈范式、纳什均衡、囚徒困境、Matching Pennies、零和博弈、混合策略
  - 姚期智不等式、NOR 估值问题 Las Vegas 时间复杂度下界

4. 球和箱子模型

- Bernoulli 分布、几何分布、二项分布
- 桶排序及其时间复杂度分析
- 跳表及其分析
- 球和箱子模型
  - 模型概述
  - 生日悖论（单射）
  - 赠券收集（满射）
  - 占用问题与负载均衡
- 生日悖论的应用
  - 生日攻击
  - Leader 选举
- 通用散列函数
  - 独立性由强到弱：相互独立、k-独立、两两独立
  - 通用散列函数、与球和箱子模型的关系
    - 2-通用散列函数族
    - 2-强通用散列函数族
- 综合运用
  - 散列表及其分析
  - 数字指纹与 Bloom Filter

5. Chernoff 界

- Chernoff 界及其常用形式
  - 矩生成函数
  - Chernoff 界的导出

    - Chernoff 界的常用形式
- 简单应用
    - 算法重复遍数（比 Markov 不等式更加精确）
    - 参数估计（基因突变概率、Π 估计、定积分估计）
- 特殊情况下更好的 Chernoff 界
- 集合平衡配置
- 超立方体上的随机路由算法
    - 超立方体拓扑结构
    - Bit Fixing 路由算法
    - 随机路由算法

6. 鞅

- 鞅的定义和基本性质
    - 条件概率、条件期望
    - 公平赌博、硬币正反面次数之差、Polya 壶
    - 鞅尾不等式、Azuma 不等式
- 鞅的一般形式
    - 随机变量的和、和的平方、杜比鞅、鞅的性质
    - 停时定理
    - 瓦尔德方程
        - 两轮骰子赌局
        - Las Vegas 算法的期望运行时间
        - 共享总线服务器通信
        - Azuma-Hoeffding 不等式
    - 鞅尾不等式随机抽样和随机舍入
- 鞅的应用
    - 模式匹配
    - 球和箱子模型中空箱子个数
    - 随机图的色数

7. 随机抽样和随机舍入

- 随机游走
    - SAT 问题的随机赋值算法分析
    - 马尔科夫链
    - 图上的随机游走
        - 二分图上的随机游走
        - 随机游走的稳定分布
        - Hitting Time、H(u, v)
- 随机抽样
    - 搜寻非二次剩余

        - ◆ 费马小定理、二次剩余和非二次剩余一样多
        - ◆ 性能分析
    - ■ 水库抽样
        - ◆ 均匀抽样问题、选择抽样算法、水库抽样算法
- 蒙特卡罗方法
    - ■ Π的计算
    - ■ DNF 满足性赋值（使一个子句满足的赋值个数）的近似计数
        - ◆ 朴素算法
        - ◆ 改造样本空间后的 Buboly-Karp 算法
    - ■ 从近似抽样到近似计数
        - ◆ 抽样的近似性
        - ◆ 近似抽样的可用性示例
    - ■ 马尔科夫链-蒙特卡罗方法
        - ◆ 独立集均匀抽样
        - ◆ Metropolis 算法
- 随机舍入
    - ■ 基本框架：松弛、小数解、整数解、近似
    - ■ 顶点覆盖问题
    - ■ 集合覆盖问题
- 混合随机算法
    - ■ MAX-SAT（CNF 最大化同时满足子句个数）的随机抽样算法 RandSample
    - ■ MAX-SAT 的随机舍入算法 RandRound
    - ■ MAX-SAT 的随机混合算法 RandMix

8. 概率方法与去随机化

- 概率论证法
    - ■ 基本计数论证
        - ◆ 坏事件不发生：同色K_k子图不出现
    - ■ 期望论证
        - ◆ 最大割问题、独立集大小、随机图的围长
    - ■ 二阶矩方法
        - ◆ General Moment Method
    - ■ Lovasz 局部引理
        - ◆ 直观含义：坏事件都不发生，独立性
        - ◆ 依赖图
        - ◆ Lovasz 引理的证明
        - ◆ 环着色、超图着色、k-SAT 可满足性
- 去随机化
    - ■ MAX-SAT 问题随机算法的去随机化
        - ◆ RandSample 去随机化
        - ◆ RandRound 去随机化

        - RandMix 去随机化
    - 集合平衡配置随机算法的去随机化
    - 随机电路去随机化

9. 代数指纹技术

- 代数指纹技术
    - 矩阵比对
    - 多项式比对
    - 位串比对
    - 字符串精确比对
- 代数指纹技术与交互式证明系统
    - 交互式证明系统
    - 图不同构问题
    - 3-SAT 问题