



華東師範大學

EAST CHINA NORMAL UNIVERSITY

数据科学与工程算法基础

Algorithm Foundations of Data Science and Engineering

第三章 尾概率不等式及其应用

$$(1+x)^n = 1 + \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} + \dots$$

课程提纲

Content

1 算法引入

2 尾概率不等式

3 Morris 算法

课程提纲

Content

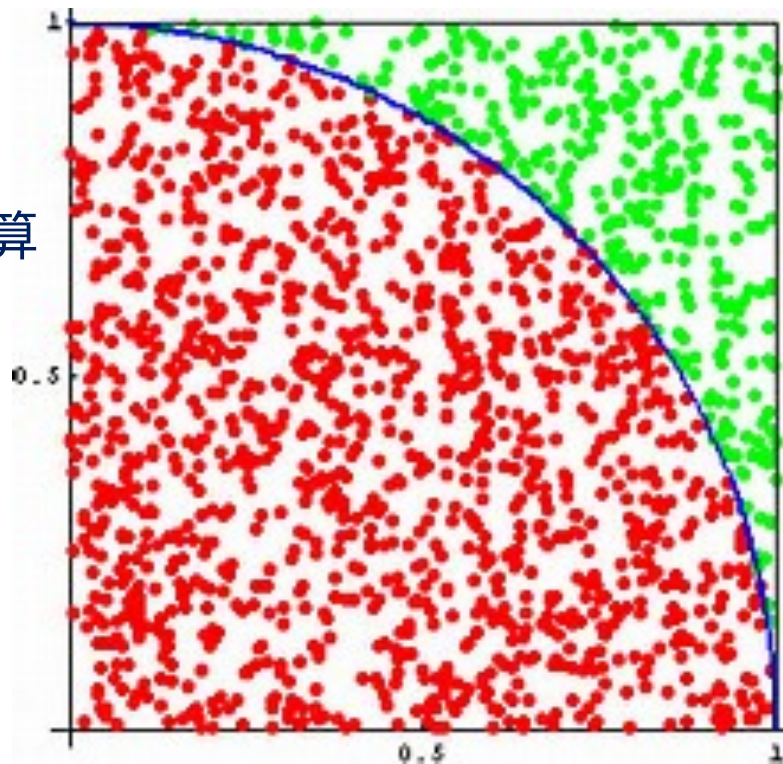
1 算法引入

2 尾概率不等式

3 Morris 算法

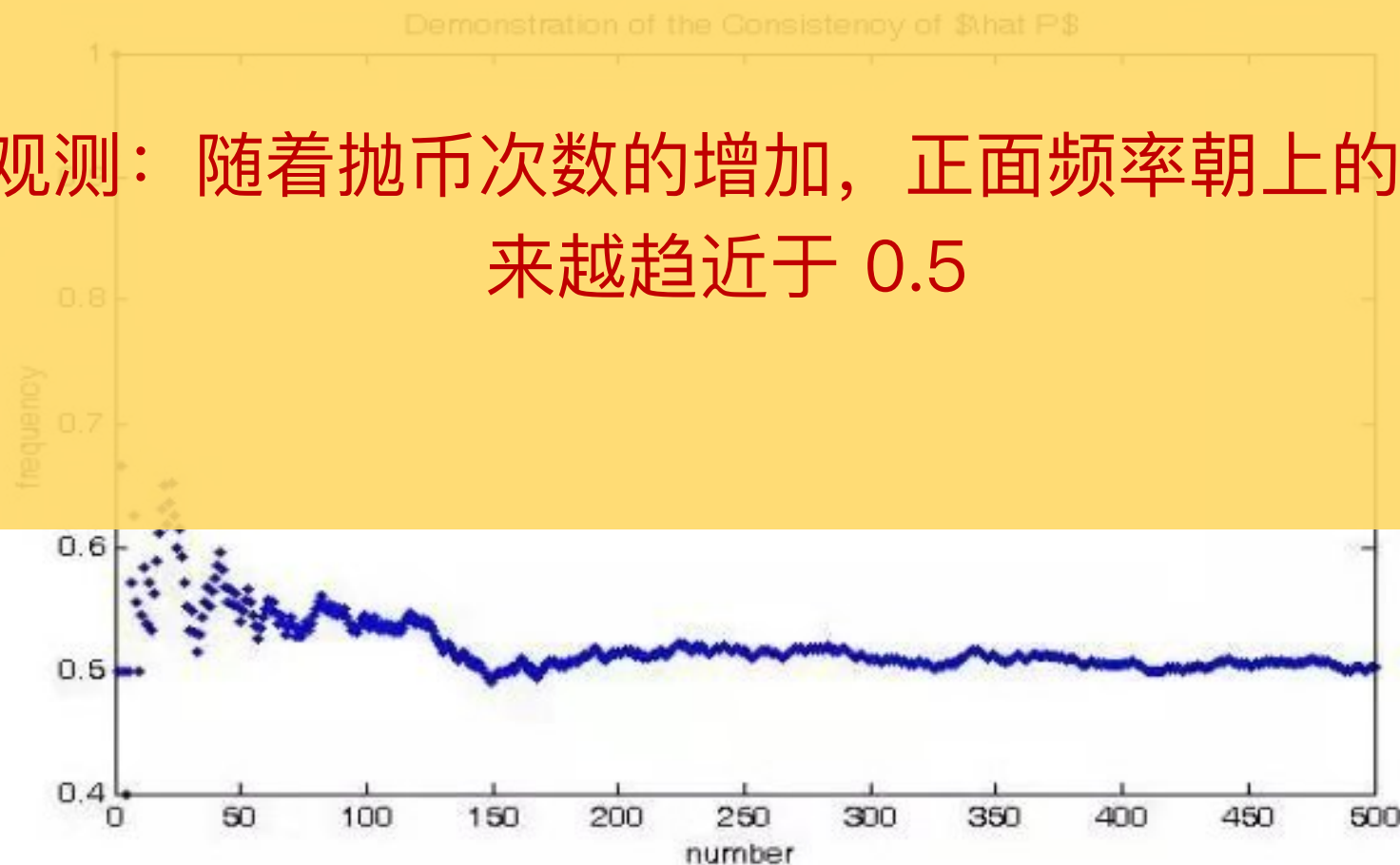
蒙特卡洛方法近似计算 π

- 蒙特卡洛(Monte Carlo)方法, 又称随机抽样或统计试验方法
 - 它是利用随机试验求解问题的方法
 - 广泛应用于计算机科学、金融工程、计算物理学等领域
- 蒙特卡洛方法近似计算 π
 - 在正方形区域内随机打点
 - 随机点落在红色区域的概率为 $\frac{\pi}{4}$
- 为什么这种方法是可行的?
 - 随机打点 1000 次, 能精确到什么程度?
 - 精确到小数点后 10 位, 至少需要打多少个点?



抛币实验

- 抛掷一枚均匀的硬币，正面朝上的频率



重要观测：随着抛币次数的增加，正面频率朝上的频率越来越趋近于 0.5

尾概率问题

- **问题描述：** 抛掷一枚均匀硬币，当抛掷 N 次后，正面朝上的次数大于 $\frac{3}{4}N$ 的概率为多少？
- **尾概率计算**
 - 随机变量尾概率与它偏离其期望 $E(X)$ 的程度息息相关
 - 精确计算尾概率往往涉及繁琐的计算
 - 能否给出尾概率的一个合理上界？
- **尾概率不等式**
 - Markov 不等式
 - Chebyshev 不等式
 - Chernoff 不等式

课程提纲

Content

1 算法引入

2 尾概率不等式

3 Morris 算法

Markov 不等式

- **Markov 不等式**：给定样本空间 Ω 上的非负随机变量 X ，对 $\forall a > 0$ ，有

$$P(X > a) \leq \frac{E(X)}{a} \text{ 或者 } P(X > aE(X)) \leq \frac{1}{a}$$

- **证明（连续随机变量）**：

因为 $X \geq 0$ ，我们得到

$$\begin{aligned} E(X) &= \int_0^{\infty} xp(x)dx = \int_0^a xp(x)dx + \int_a^{\infty} xp(x)dx \\ &\geq \int_a^{\infty} xp(x)dx \geq a \int_a^{\infty} p(x)dx = aP(X > a) \end{aligned}$$

所以，我们得到 $P(X > a) \leq \frac{E(X)}{a}$ 。

Markov 不等式

- **Markov 不等式：** 给定样本空间 Ω 上的非负随机变量 X ，对 $\forall a > 0$ ，有

$$P(X > a) \leq \frac{E(X)}{a} \text{ 或者 } P(X > aE(X)) \leq \frac{1}{a}$$

- 例如，均匀硬币抛掷 n 次后，正面朝上次数 X 大于 $\frac{3}{4}n$ 的概率 $P(X > \frac{3}{4}n) \leq \frac{n/2}{3n/4} = \frac{2}{3}$
- 从这个例子来看，Markov不等式给出的尾概率上界可能非常宽松

例题：Markov 不等式

- 抛掷一枚正面朝上概率为 $\frac{1}{3}$ 的不均匀硬币，当抛掷 N 次后，正面朝上的次数大于 $\frac{N}{2}$ 的概率为多少？

Chebyshev 不等式

- **Chebyshev 不等式：**假定随机变量 X 的期望与方差分别为 μ 和 σ^2 ，则

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

- **证明：**令随机变量 $Y = (X - \mu)^2$ ，根据 Markov 不等式

$$P(Y > a^2) \leq \frac{E(Y)}{a^2} = \frac{\sigma^2}{a^2}$$

- 上例 $P(X > \frac{3}{4}n) < P(|X - \frac{n}{2}| > \frac{n}{4}) \leq \frac{\text{Var}(X)}{(n/4)^2} = \frac{16np(1-p)}{n^2} = \frac{4}{n}$
- 如果抛掷均匀硬币 1000 次，这一尾概率小于 0.004

例题：Chebyshev不等式

- 抛掷一枚正面朝上概率为 $\frac{1}{3}$ 的不均匀硬币，当抛掷 N 次后，正面朝上的次数大于 $\frac{N}{2}$ 的概率为多少？

$b(n, \frac{1}{3})$ 不是对称分布!

Chernoff 不等式

- **Chernoff 不等式 (下界)** : 设 X_i 是独立的伯努利随机向量序列, 其中 $P(X_i = 1) = p_i$, 令 $X = \sum_{i=1}^n X_i$ 和 $\mu = \sum_{i=1}^n p_i$, 对任意 $\delta \in (0,1)$, 则

$$1) P(X < (1 - \delta)\mu) < \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu$$

$$2) P(X < (1 - \delta)\mu) < \exp\left(-\frac{\mu\delta^2}{2}\right)$$

- **证明 (结论 1)** : 对任意 $t > 0$, 我们有

$$X < (1 - \delta)\mu \Rightarrow tX < t(1 - \delta)\mu \Rightarrow e^{tX} < e^{t(1-\delta)\mu} \Rightarrow e^{-tX} > e^{-t(1-\delta)\mu}$$

因此, 我们得到 $P(X < (1 - \delta)\mu) = P(\exp(-tX) > \exp(-t(1 - \delta)\mu))$

进一步的, 因为 $\exp(-tX) = \exp\left(-t \sum_{i=1}^n X_i\right) = \prod_{i=1}^n \exp(-tX_i)$,

我们得到 $P(X < (1 - \delta)\mu) < \frac{\prod_{i=1}^n E(\exp(-tX_i))}{\exp(-t(1 - \delta)\mu)}$ (由Markov不等式)

Chernoff 不等式 (证明续)

由于 $1 - x < e^{-x}$, 所以

$$E(\exp(-tX_i)) = p_i e^{-t} + (1 - p_i)e^0 = 1 - p_i(1 - e^{-t}) < \exp(p_i(e^{-t} - 1))$$

进一步计算, 我们得到

$$\prod_{i=1}^n E(\exp(-tX_i)) < \prod_{i=1}^n \exp(p_i(e^{-t} - 1)) = \exp(\mu(e^{-t} - 1))$$

$$\text{因此, } P(X < (1 - \delta)\mu) \leq \frac{\exp(\mu(e^{-t} - 1))}{\exp(-t(1 - \delta)\mu)} = \exp(\mu(e^{-t} + t - t\delta - 1))$$

选择 t 值得到更为紧致的上界, 对上式指数求导并令其为零, 即令 $-e^{-t} + 1 - \delta = 0$, 就得到 $t = \ln(1/(1 - \delta))$, 将其代入得

$$P(X < (1 - \delta)\mu) < \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu$$

Chernoff 不等式 (证明续)

- 为得到一个更简洁的上界, 需要丢掉冗杂的 $(1 - \delta)^{1-\delta}$
- 基于泰勒展示式, 易知

$$(1 - \delta)\ln(1 - \delta) = (1 - \delta) \cdot \sum_{i=1}^{\infty} \left(-\frac{\delta^i}{i}\right) > -\delta + \frac{\delta^2}{2}$$

- 因此,

$$P(X < (1 - \delta)\mu) \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^{\mu} < \left(\frac{e^{-\delta}}{\exp(-\delta + \frac{\delta^2}{2})}\right)^{\mu} = \exp\left(-\frac{\mu\delta^2}{2}\right)$$

Chernoff 不等式 (续)

- **Chernoff 不等式 (上界)**：设 X_i 是独立的伯努利随机向量序列且 $P(X_i = 1) = p_i$ ，令 $X = \sum_{i=1}^n X_i$ 和 $\mu = \sum_{i=1}^n p_i$ ，对任意 $\delta \in (0,1)$ ，则

$$1) P(X > (1 + \delta)\mu) < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu$$

$$2) P(X > (1 + \delta)\mu) < \exp\left(-\frac{\mu\delta^2}{2 + \delta} \right)$$

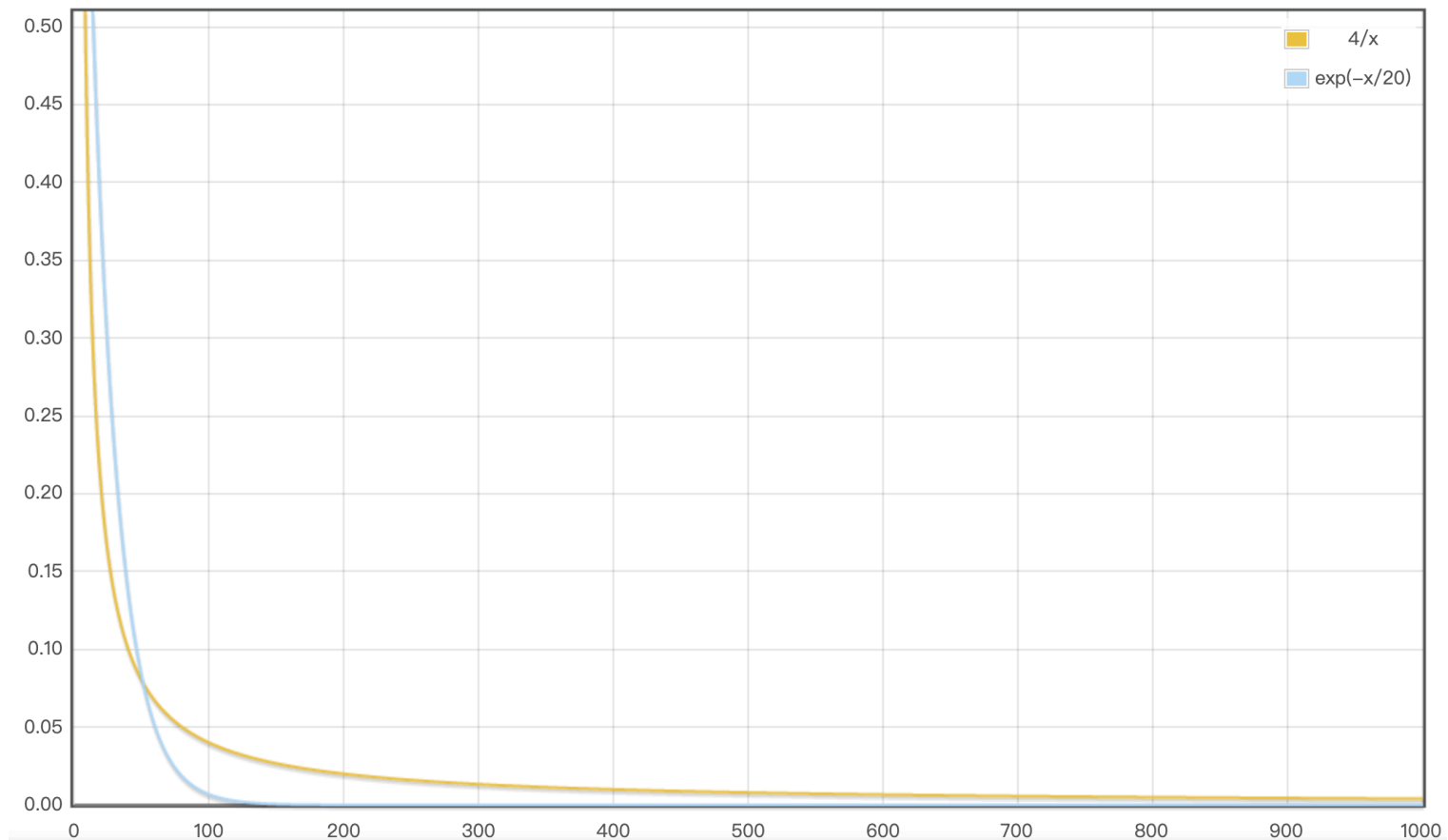
$$3) P(X > (1 + \delta)\mu) < \exp\left(-\frac{\mu\delta^2}{3} \right)$$

- **证明**：公式 1 的证明和下界类似，留作课后练习题。公式 2 的证明需要对公式 1 右侧取对数，得到 $\mu(\delta - (1 + \delta)\ln(1 + \delta))$ 。当 $\delta > 0$ ，我们有 $\ln(1 + \delta) \geq \frac{\delta}{1 + \delta/2}$ (留作课后练习题)，由此得到公式 2。最后，公式 3 是公式 2 在 $\delta \in (0,1)$ 时的简化形式。

Chernoff 不等式 (续)

- **Chernoff 不等式 (上界)** : 设 X_i 是独立的伯努利随机向量序列且 $P(X_i = 1) = p_i$, 令 $X = \sum_{i=1}^n X_i$ 和 $\mu = \sum_{i=1}^n p_i$, 对任意 $\delta \in (0,1)$, 则
 - 1) $P(X > (1 + \delta)\mu) < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu$
 - 2) $P(X > (1 + \delta)\mu) < \exp\left(-\frac{\mu\delta^2}{2 + \delta} \right)$
 - 3) $P(X > (1 + \delta)\mu) < \exp\left(-\frac{\mu\delta^2}{3} \right)$
- 上例 $P(X > \frac{3n}{4}) = P(X > (1 + \frac{1}{2})\frac{n}{2}) < \exp(-\frac{n}{2}(\frac{1}{2})^2/3) = \exp(-\frac{n}{24})$
- 如果抛掷均匀硬币 1000 次, 这一尾概率小于 $\exp(-1000/24) \approx 8^{-19}$

Chebyshev和Chernoff的上界比较



例题：Chernoff 不等式

- 抛掷一枚正面朝上概率为 $\frac{1}{3}$ 的不均匀硬币，当抛掷 N 次后，正面朝上的次数大于 $\frac{N}{2}$ 的概率为多少？

扩展阅读：Hoeffding不等式

Let X_1, \dots, X_n be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely. Consider the sum of these random variables,

$$S_n = X_1 + \dots + X_n.$$

Then Hoeffding's theorem states that, for all $t > 0$,^[3]

$$P(S_n - E[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

$$P(|S_n - E[S_n]| \geq t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Here $E[S_n]$ is the expected value of S_n .

其他经典概率不等式

- Azuma不等式
- Bennett不等式
- Bernstein不等式
-

课程提纲

Content

1 算法引入

2 尾概率不等式

3 Morris 算法

计数问题

- 假设有一台路由器，主要负责转发网络流量包
 - 其中包括源地址 IP 和目标地址 IP 等信息
- 统计指向某个 IP 地址的请求数量
 - 可能请求数量很多
 - 但是路由器的存储容量有限
- 因此，精确计数空间消耗过大
 - 可以给出近似计数
 - 但是空间消耗更小
- 近似计数方法
 - Morris 算法
 - Morris+ 算法
 - Morris++ 算法



Morris 算法

- Morris 算法

- 初始化 $X \leftarrow 0$
- 对每一次事件发生
 - 变量 X 的值以 $\frac{1}{2^X}$ 的概率增 1
- 对每次查询，输出 $\hat{n} = 2^X - 1$

- 算法特点

- 监视某个事件是否发生
- 实时返回当前的计数值大小
- 不直接估计计数值的大小，而是估计保存二进制计数值所需位数
- 因此，空间复杂度为 $O(\log_2 n)$ ，其中 n 为真实计数值

Morris 算法示例

input	True	Sampling probability	X	Estimator
	0	1	0	0
1	1	$\frac{1}{2}$	1	1
1	2	$\frac{1}{2}$	1	1
1	3	$\frac{1}{4}$	2	3
1	4	$\frac{1}{4}$	2	3
1	5	$\frac{1}{4}$	2	3
1	6	$\frac{1}{4}$	2	3
1	7	$\frac{1}{8}$	3	7
1	8	$\frac{1}{8}$	3	7

通过这个例子，你能发现 Morris 算法的缺点吗？

Table 1: Running example of Morris

Morris 算法的无偏性

- **定理：**假定 n 次更新后， X_n 表示 Morris 算法的计数器 X ，则 $E(2^{X_n} - 1) = n$
- **证明：**

当 $X_n = j$ 时， $X_{n+1} = j + 1$ 的概率为 2^{-j} ，而 $X_{n+1} = j$ 的概率为 $1 - 2^{-j}$ 。根据条件概率的性质得到 $E(2^{X_{n+1}}) = \sum_{j=1}^{\infty} P(X_n = j)E(2^{X_{n+1}} | X_n = j)$,

其中 $E(2^{X_{n+1}} | X_n = j) = 2^{j+1}2^{-j} + 2^j(1 - 2^{-j}) = 2^j + 1$ 。最后，我们使用数学归纳法得到结论。

Morris 算法的无偏性

- **定理：**假定 n 次更新后， X_n 表示 Morris 算法的计数器 X ，则 $E(2^{X_n} - 1) = n$
- **证明（续）：**

当 $n = 1$ 时， $X_1 = 1$ 的概率为 1。因此， $E(2^{X_1} - 1) = 2^1 - 1 = 1$ 结论成立。

假设当 $n = k$ 时， $E(2^{X_k} - 1) = k$ 成立，则当 $n = k + 1$ 时，

$$\begin{aligned} E(2^{X_{k+1}}) &= \sum_{j=1}^{\infty} P(X_k = j) E(2^{X_{k+1}} | X_k = j) = \sum_{j=1}^{\infty} P(X_k = j) (2^j + 1) \\ &= \sum_{j=1}^{\infty} P(X_k = j) (2^j - 1) + \sum_{j=1}^{\infty} P(X_k = j) 2 \\ &= E(2^{X_k} - 1) + 2 = k + 2 \end{aligned}$$

结论 $E(2^{X_n} - 1) = n$ 得证。

Morris 算法的方差

- X_n 表示 n 次更新后 Morris 算法中的计数器 X , 则 $\text{Var}(2^{X_n}) = O(n^2)$ 。
- 证明: 首先, 我们有

$$\text{Var}(2^{X_n}) = E(2^{2X_n}) - E^2(2^{X_n}) = E(2^{2X_n}) - (n+1)^2$$

$$\begin{aligned} E(2^{2X_n}) &= \sum_{i=1}^{\infty} 2^{2i} P(X_n = i) = \sum_{i=1}^{\infty} 2^{2i} \left(\frac{1}{2^{i-1}} P(X_{n-1} = i-1) + \left(1 - \frac{1}{2^{i-1}}\right) P(X_{n-1} = i) \right) \\ &= \sum_{i=1}^{\infty} 2^{i+1} P(X_{n-1} = i-1) + \sum_{i=1}^{\infty} 2^{2i} P(X_{n-1} = i) - \sum_{i=1}^{\infty} 2^i P(X_{n-1} = i) \\ &= E(2^{2X_{n-1}}) + 3 \sum_{i=1}^{\infty} 2^i P(X_{n-1} = i) = E(2^{2X_{n-1}}) + 3E(2^{X_{n-1}}) \end{aligned}$$

$$\text{已知 } E(2^{2X_0}) = 1, \text{ 因此 } E(2^{2X_n}) = 3 \sum_{i=0}^{n-1} E(2^{X_i}) + 1 = \frac{3n(n+1)}{2} + 1$$

$$\text{所以, } \text{Var}(2^{X_n}) \approx \frac{n^2}{2} = O(n^2)$$

Morris 算法分析

- 令输出 $\hat{n} = 2^X - 1$ ，根据 Chebyshev 不等式，

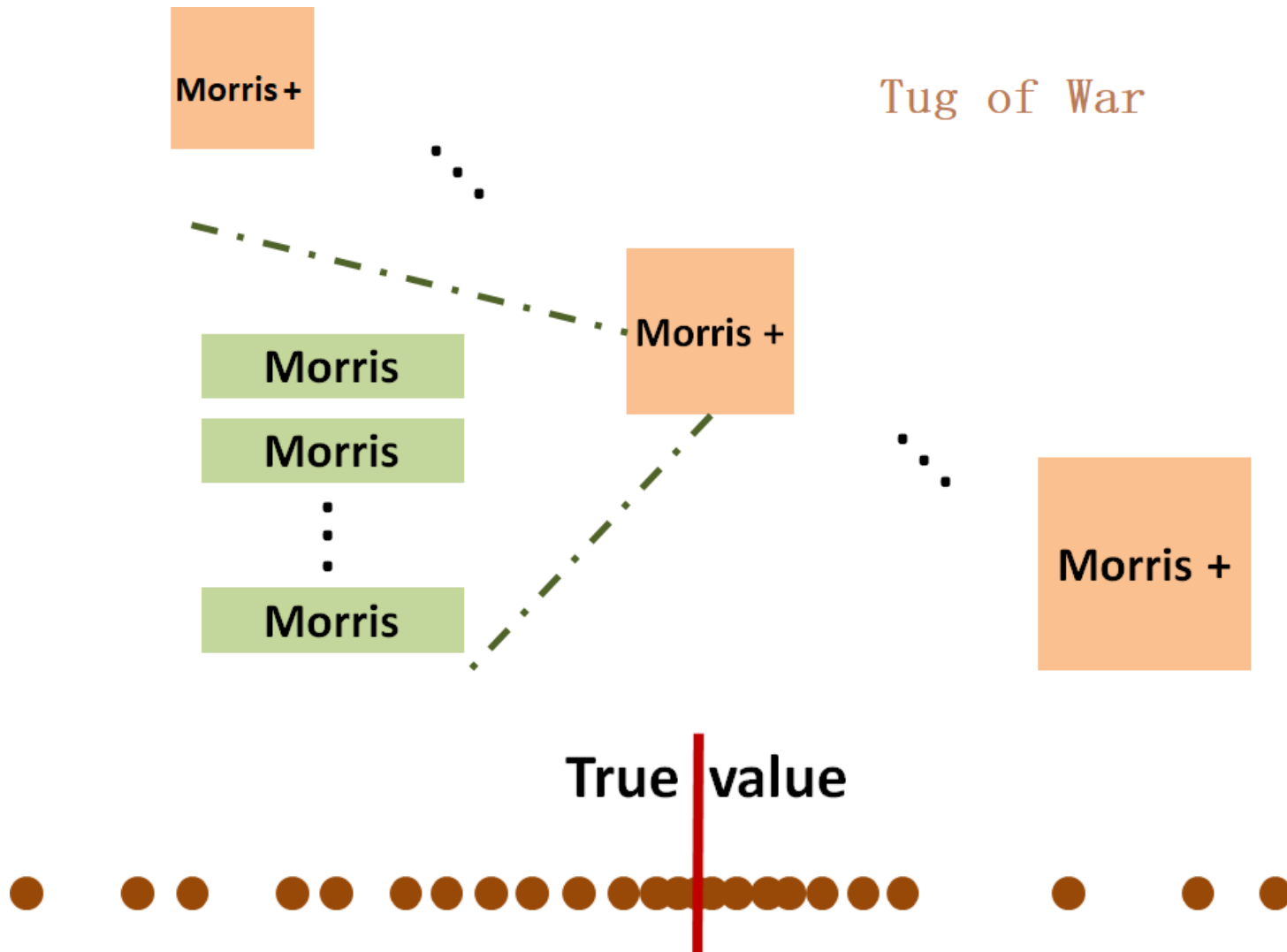
$$P(|\hat{n} - n| > \epsilon n) < \frac{n^2}{2\epsilon^2 n^2} = \frac{1}{2\epsilon^2}$$

- 结果表明，仅当 $\epsilon \geq 1$ 时，尾概率上界不大于 $1/2$
- Morris 算法优缺点
 - 输出结果是真实值的无偏估计
 - 但是可能偏离真实值过大
 - 需要进一步降低算法输出结果的误差
- 如果对于一个很小的 ϵ ，输出结果满足 $|\hat{n} - n| \leq \epsilon n$ 视为成功，那么如何才能大概率保证算法输出成功输出理想的结果呢？

Morris+ 算法

- 为降低 Morris 算法输出结果的方差, Morris+ 算法
 - 运行 k 个独立的 Morris 算法得到 k 个不同的计数值 X_1, \dots, X_k
 - 对每次查询, 输出平均值 $\hat{n} = \frac{1}{k} \sum_{i=1}^k 2^{X_i} - 1$
- Morris+ 算法
 - k 个独立的 Morris 算法结果分别为 $\hat{n}_1, \dots, \hat{n}_k$, 输出 $\hat{n} = \frac{1}{k} \sum_{i=1}^k \hat{n}_i$
 - 根据 Chebyshev 不等式, $P(|\hat{n} - n| > \epsilon n) < \frac{O(n^2/k)}{\epsilon^2 n^2} \approx \frac{1}{k\epsilon^2} < \delta$
 - 当 $k > \frac{1}{\delta\epsilon^2}$, $P(|\hat{n} - n| > \epsilon n) < \delta$
 - 因此, Morris+ 算法复杂度为 $O(\frac{1}{\delta\epsilon^2})$

Morris++ 算法



算法细节

- 运行 t 个 Morris+ 实例，每一个都有 $1/3$ 的失败概率，即

$$P(|\hat{n} - n| > \epsilon n) < \frac{1}{k\epsilon^2} = \frac{1}{3}, \text{ 其中 } k = O\left(\frac{1}{\epsilon^2}\right)$$

- 输出所有 t 个 Morris+ 实例输出值的中位数
 - t 个 Morris+ 算法运行失败的期望不超过 $\frac{t}{3}$
 - 如果 Morris++ 算法的输出是一个糟糕的估计，则意味着 t 个 Morris+ 算法至少有一半失败了
- 这种方法被称之为 Tug of War
 - 可以将算法复杂度降为 $O\left(\frac{\ln(1/\delta)}{\epsilon^2}\right)$
 - 比 Morris+ 算法更为高效



Morris++ 算法分析

- 定义随机变量 $Y_i = \begin{cases} 1, & \text{if } |\widehat{n}_i - n| > \epsilon n \\ 0, & \text{otherwise} \end{cases}$
 - 随机变量 Y_i 表示第 i 次 Morris+ 算法失败与否
 - 令 $k = O(\frac{1}{\epsilon^2})$, 每次 Morris+ 算法运行失败的概率不超过 $1/3$

- 注意到, $\mu = E(\sum_i Y_i) \leq \frac{t}{3}$, 则根据 Chernoff 不等式可得

$$P(\sum_i Y_i > \frac{t}{2}) \leq P(\sum_i Y_i > (1 + \frac{1}{2})\mu) \leq \exp(-\mu(\frac{1}{2})^2/3) < \delta$$

也就是说, $\mu < 12 \ln(1/\delta)$

- 因此, 需要运行 $t = O(\ln(1/\delta))$ 次 Morris+ 算法, 其结果的中位数是符合精度要求的近似估计
- Morris++ 算法复杂度为 $O(\frac{\ln(1/\delta)}{\epsilon^2})$

本章小结

- 尾概率不等式
 - Markov 不等式
 - Chebyshev 不等式
 - Chernoff 不等式
 -
- 尾概率不等式的应用
 - 流数据分析
 - 近似算法分析
 - 大偏差理论
 -

课后作业

课本第45–46页习题3

- 第1, 2, 3, 4, 5, 6题
- 第7题难度较大, 仅供大家课后思考不做考察