Advanced Algorithmics

Strategies for Tackling Hard Problems

Sebastian Wild

Markus Nebel

Lecture 10

2017-05-22

4.2 Classification of Randomized Algorithms

Consider here the general problem to compute some *function* $f : \Sigma^\star \to \Gamma^\star$.

$\rightsquigarrow$ Covers *decision problems* $L \subseteq \Sigma^\star$ by setting $\Gamma = \{0,1\}$ and $f(w) = \begin{cases} 1 & w \in L \\ 0 & w \notin L \end{cases}$

Definition 4.1 (Las Vegas Algorithm)
A randomized algorithm A is a *Las-Vegas (LV) algorithm* for a problem $f : \Sigma^\star \to \Gamma^\star$ if for all $x \in \Sigma^\star$ holds

1. $\Pr\big[time_A(x) < \infty\big] = 1$ (*finite* number of computations)
2. $A(x) \in \{f(x), ?\}$ (answer always *correct or "don't know"*)
3. $\Pr\big[A(x) = f(x)\big] \geq \frac{1}{2}$ (*correct half the time*)

Theorem 4.2 (Don't know don't needed)

Every Las Vegas algorithm A for $f : \Sigma^\star \to \Gamma^\star$ can be transformed into a randomized algorithm B for f so that for all $x \in \Sigma^\star$ holds

1. $\Pr[B(x) = f(x)] \;=\; 1$ (always correct)
2. $\mathbb{E}\text{-}time_B(x) \;\le\; 2 \cdot time_A(x)$

◀

Theorem 4.3 (Termination Enforcible)

Every randomized algorithm B for $f : \Sigma^\star \to \Gamma^\star$ with $\Pr[B(x) = f(x)] = 1$ can be transformed into a Las Vegas algorithm A for f so that for all $x \in \Sigma^\star$ holds

$$time_A(x) \;\le\; 2 \cdot \mathbb{E}\text{-}time_B(x).$$

◀

$\rightsquigarrow$ *Can trade expected time bound for worst-case bound by allowing "don't know" and vice versa!*

Both types are called LV algorithms.

Las Vegas Examples

rollDie by rejection sampling is Las Vegas of unbounded worst-case type.

Easy to transform into Las Vegas according to Definition 4.1:

---

procedure rollDieLasVegas:
    Draw 3 random bits b_2, b_1, b_0
    $n = \sum_{i=0}^{2} 2^i b_i$ *// Interpret as binary representation of a number in* $[0:7]$
    if $(n = 0 \lor n = 7)$
        return $?$
    else
        return n

---

randomized

Other famous examples: *Quicksort* and *Quickselect*

- always correct *and*
- $time(n) = \mathcal{O}(n^2) < \infty$
- much better average:
  - $\mathbb{E}\text{-}time_{QSort}(n) = \Theta(n \log n)$
  - $\mathbb{E}\text{-}time_{QSelect}(n) = \Theta(n)$

$\mathbb{E}$ over inputs: Quicksort with pivot say $A[0]$ on random permutations $\sim 2n\ln n$

$\mathbb{E}$ over random bits: Quicksort with random pivot say $A[P]$ $\sim 2n\ln n$

$\boxed{\text{input does not matter!}}$

To Err is Algorithmic

Sometimes sensible to allow *wrong/imprecise* answers ... but random should not mean *arbitrary*.

Definition 4.4 (Monte Carlo Algorithm)

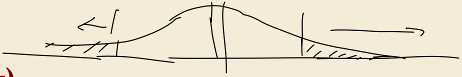
A randomized algorithm A is a *Monte Carlo algorithm* for $f : \Sigma^\star \to \Gamma^\star$

- with *bounded error* if $\quad \exists \varepsilon > 0 \, \forall x \in \Sigma^\star \quad : \quad \Pr[A(x) = f(x)] \;\geq\; \dfrac{1}{2} + \varepsilon.$
- with ***un**bounded error* if $\quad \forall x \in \Sigma^\star \quad : \quad \Pr[A(x) = f(x)] \;>\; \dfrac{1}{2}.$

$\varepsilon = \varepsilon(x)$

Seems like a minuscule difference? We will see it is vital!

4.3 Tail Bounds and Concentration of Measure



Theorem 4.5 (Markov's Inequality)

Let $X \in \mathbb{R}_{\geq 0}$ be a r.v. that assumes only *weakly positive* values. Then holds

$$\forall a > 0 \quad : \quad \Pr[X \geq a] \;\leq\; \frac{\mathbb{E}[X]}{a}$$

Proof: Let $a > 0$ define $I = \mathbb{1}_{\{X \geq a\}} = [X \geq a] = \begin{cases} 1 & X \geq a \\ 0 & \text{else} \end{cases}$

$I \leq \frac{X}{a} \mid \mathbb{E}$

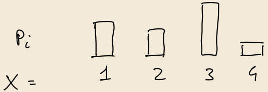
- $X < a \Rightarrow I = 0$ but $X, a \geq 0$
- $X \geq a \Rightarrow I = 1$ $\quad \frac{X}{a} \geq 1$

$$\Pr[X \geq a] = \mathbb{E}[I] \leq \mathbb{E}\left[\frac{X}{a}\right] = \frac{\mathbb{E}[X]}{a} \qquad \square$$

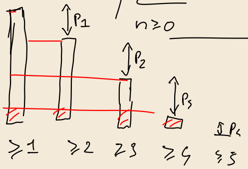
Since $X \geq 0$ implies $\mathbb{E}X \geq 0$, nicer equivalent form: $\boxed{\forall a > 0 : \Pr\big[X \geq a\mathbb{E}[X]\big] \;\leq\; \frac{1}{a}}$

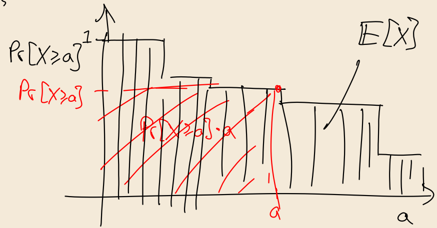
Markov Mnemonic:

If $X \in \mathbb{N}_0$ r.v. $\Rightarrow$ $\mathbb{E}[X] = \sum_{n \geq 0} n \cdot \Pr[X=n] = \boxed{\sum_{n \geq 0} \Pr[X \geq n] = \mathbb{E}[X]}$



$1 \cdot p_1 + 2 \cdot p_2 + 3 p_3 + 4 p_4$





$\mathbb{E}[X \geq a) \cdot a \leq \mathbb{E}[X]$

Markov's inequality is often giving weak bounds, but can often apply if some $f(X) \geq 0$

Definition 4.6 (Moments, variance, standard deviation)

For random variable X, $\mathbb{E}[X^k]$ is the *kth moment* of X. 1st $\mathbb{E}[X]$

The *variance* (second centered moment) of X is given by $\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ and its *standard deviation* is $\sigma[X] = \sqrt{\mathrm{Var}[X]}$.

Theorem 4.7 (Chebychev's Inequality)

Let X be a random variable. We have

$$\forall a > 0 \quad : \quad \Pr\left[|X - \mathbb{E}[X]| \geq a\right] \;\leq\; \frac{\mathrm{Var}[X]}{a^2}$$

Proof:

$$\Pr[\,|X - \mathbb{E}[X]| \geq a\,] = \Pr[\,\overbrace{(X - \mathbb{E}[X])^2}^{\geq 0} \geq a^2\,]$$

$$\underset{\text{Markov}}{\leq} \frac{\mathbb{E}[(X - \mathbb{E}X)^2]}{a^2} = \frac{\mathrm{Var}[X]}{a^2} \qquad \square$$

"Trick": Centering $(-\mathbb{E}[X])$ and taking power made variable "more variable" → stronger bound from Markov.

Corollary 4.8 (Chebychev Concentration)

Let $X_1, X_2, \ldots$ be a sequence of random variables and assume

- $\mathbb{E}[X_n]$ and $\mathrm{Var}[X_n]$ exist for all n and
- $\sigma[X_n] = o(\mathbb{E}[X_n])$ as $n \to \infty$.

Then holds

$$\forall \varepsilon > 0 \quad : \quad \Pr\left[\left|\frac{X_n}{\mathbb{E}[X_n]} - 1\right| \geq \varepsilon\right] \to 0 \qquad (n \to \infty),$$

i.e., $\frac{X_n}{\mathbb{E}[X_n]}$ *converges in probability* to 1.

Chernoff Bounds

For specific distribution, much stronger tail concentration inequalities are possible.

Theorem 4.9 (Chernoff Bound for Poisson trials)

$p_1 = \cdots = p_n$ $\rightsquigarrow$ Bernoulli trials

Let $X_1, \ldots, X_n \in \{0,1\}$ be *(mutually) independent* with $X_i \overset{\mathcal{D}}{=} B(p_i)$. Define $X = X_1 + \cdots + X_n$ and $\underbrace{\mu}_{= \mathbb{E}[X]} = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = p_1 + \cdots + p_n$. Then holds

$$\forall \delta > 0 \quad : \quad \Pr[X \geq (1+\delta)\mu] \; < \; \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

$$\forall \delta \in (0,1] \quad : \quad \Pr[X \geq (1+\delta)\mu] \; \leq \; \exp(-\mu\delta^2/3)$$

follows

Proof: Let $t > 0$

$$\Pr[X \geq (1+\delta)\mu] = \Pr\left[\overbrace{e^{tX}}^{\geq 0} \geq e^{t(1+\delta)\mu}\right]$$

$$\underset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}}$$

$$= \frac{1}{e^{t(1+\delta)\mu}} \cdot \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} tX_i\right)\right]$$

$$= \quad '' \quad \cdot \mathbb{E}\left[\prod_{i=1}^{n} \exp(tX_i)\right]$$

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

mut. indep.

$$= \prod_{i} \left(p_i e^t + (1-p_i)\,1\right)$$

$$= \prod_i \left(1 + p_i\left(e^t - 1\right)\right)$$

$$\leq \prod_i e^{p_i(e^t - 1)}$$

$$= \exp\Big(\left(e^t - 1\right)\overbrace{\sum_i p_i}^{\mu}\Big)$$

$$= \frac{e^{\mu(e^t - 1)}}{e^{t(1+\delta)\mu}}$$

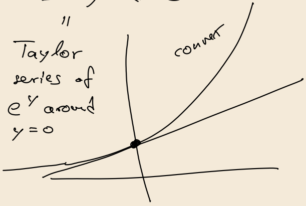
$$= \frac{e^{\mu(1+\delta-1)}}{(1+\delta)^{(1+\delta)\mu}}$$

$$= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

"convex function lies above its tangents"

$$1 + y \leq e^y$$

" Taylor series of e^y around $y=0$



$\forall\, t > 0$

good choice $t = \ln(1+\delta) > 0$

$\square$

Corollary 4.10 (Chernoff Bound for Binomial Distribution)

Let $X \overset{\mathcal{D}}{=} \underline{\mathrm{Bin}(n,p)}$. Then $\mathbb{E}[X] = n \cdot p$

$X = X_1 + \dots + X_n$

$X_i \overset{\mathcal{D}}{=} \mathcal{B}(p)$

$$\forall \delta \geq 0 \quad : \quad \Pr\left[\left|\frac{X}{n} - p\right| \geq \delta\right] \leq 2\exp(-2\delta^2 n)$$

Application 1: Can we trust Quicksort's expectation?

Definition 4.11 (With high probability)

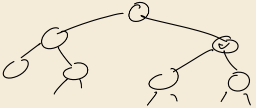
We say

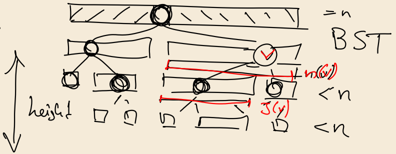
- an event $X = X(n)$ happens *with high probability (w.h.p.)* when
  $\forall c \ : \ \Pr[X(n)] \ = \ 1 \pm \mathcal{O}(n^{-c})$ as $n \to \infty$.
- a random variable $X = X(n)$ is *in* $\mathcal{O}(f(n))$ *with high probability (w.h.p.)* when
  $\forall c \exists d \ : \ \Pr[X \le df(n)] \ = \ 1 \pm \mathcal{O}(n^{-c})$ as $n \to \infty$.
  (This means, the constant in $\mathcal{O}(f(n))$ may depend on c.)

Theorem 4.12 (Quicksort Concentration)

The height of the recursion tree of (randomized) Quicksort is in $\mathcal{O}(\log n)$ w.h.p.

Hence the number of comparisons are in $\mathcal{O}(n \log n)$ w.h.p.

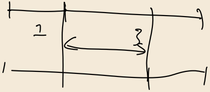




Proof: v : node in recursion tree

$n(v)$: #elems in the subtree of v

$J(v)$: size of the left child

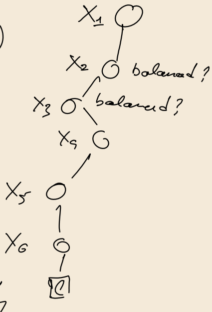


v balanced $\Leftrightarrow$ $n(v) \leq 1 \ \vee\ \frac{1}{4} \leq \frac{J(v)}{n(v)} \leq \frac{3}{4}$

↳ reduces subtree size of its child to $\leq \frac{3}{4} n(v)$

(*) Any recursion tree for n elements can contain at most $\log_{3/4}(1/n) = \log_{4/3}(n) \leq 3.5 \ln(n)$ balanced nodes. :

$$n \cdot \left(\frac{3}{4}\right)^{\log_{3/4}(1/n)} = 1 \quad \therefore (*)$$



Problem: to apply Chernoff to $X = X_1 + \dots + X_n$

we need that $X_1, \dots, X_n$ mutually independent ↯