

RANDOMIZED ALGORITHMIC TECHNIQUES AND THEIR APPLICATIONS

RECAP :

- Randomized algorithm : uses $X = \text{RANDOM}(S)$ as a primitive : X is a random variable uniformly distributes over S .
- $X = \text{RANDOM}(S_1); Y = \text{RANDOM}(S_2)$
 $\Rightarrow X$ and Y independent:
 $\Pr(X=x, Y=y) = \Pr(X=x) \cdot \Pr(Y=y)$
- Given A , randomized algorithm,
correctness and running time are
random variables:
 - $\Pr(A \text{ is correct on instances of size } n)$
 - $\Pr(T_A(n) \geq f(n))$, where
 $T_A(n)$ is the random variable
associated to worst-case complexity
 \Leftrightarrow analysis + forte \Rightarrow general w.t. per expect.
 - $E[T_A(n)] = \Theta(g(n))$

REMARK Analysis is done for fixed input (worst case). Not to be confused with probabilistic analysis of (deterministic) algorithms where the input is drawn from some distribution (compare random prime testing vs Miller-Rabin)

RANDOMIZED ALGORITHMS: VARIANTS

- Algorithms that are always correct for each run and each input:

$$\Pr(i \in A(i)) = 1$$

These are called LAS VEGAS algorithms

Randomization only affects running time

EXAMPLE: RANDOMIZED QUICKSORT

- Algorithms that can be incorrect:

$$\exists i \in \mathbb{N}: \Pr(i \notin A(i)) > 0$$

Randomization affects correctness but may affect running time (or not)

EXAMPLE: MILLER RABIN PRIMALITY-TEST

KARGER'S MIN CUT ALGORITHM

These are called MONTE CARLO algorithms

- solutions: one-sided / two-sided error Monte Carlo algorithms for decision problems (recall MILLER-RABIN)

RANDOMIZATION IS A powerful tool:

- Often allows us to solve problems which are difficult to attack deterministically
- Randomization often provides simple and practical solutions: random choices help break worst-case scenarios (e.g. symmetry breaking in the exponential backoff ethernet protocol)
- Simplicity is often counterbalanced by increased complexity in the analysis

RUNNING TIME ANALYSIS

There are two types of analyses that can be performed on a randomized algorithm. Let $T_A(n)$ be the random variable denoting the worst case time on instances of size n :

- **ANALYSIS IN EXPECTATION**

Study $E[T_A(n)]$

- **ANALYSIS IN HIGH PROBABILITY**

We say that A_{R} has running time $T(n) = O(f(n))$ with high probability (w.h.p.) if

$\exists c, \delta > 0, n_0 \in \mathbb{N} : \forall n \geq n_0 :$

$$\Pr_{A_{\text{R}}} (T(n) \geq c \cdot f(n)) \leq \frac{1}{n^\delta}$$

• CORRECTNESS ANALYSIS

For Monte Carlo algorithms, we can study correctness w.h.p.:

$$\exists \delta > 0, \forall \epsilon > 0: \forall n \geq n_0: \Pr_{\text{rand}}[A_n(i) \neq A_\pi(i)] \leq \frac{\delta}{n^\epsilon}$$

REMARK by choosing

$$\delta = \epsilon \log |K_u|$$

HILLER-LARSEN can be made correct

w.h.p.:

$$\Pr[A_n(u) \text{ incorrect}] \leq \frac{1}{2^3} = \frac{1}{2^{\log_2 K_u}} = \frac{1}{K_u} \delta^2$$

ANALYSIS W.H.P. IS OFTEN STRONGER THAN ANALYSIS W EXPECTATION

We will use Herkos's Lemma that relates EXPECTATION to the TAIL OF THE DISTRIBUTION of a random variable

MARKOV'S LEMMA

Let T be a nonnegative integer

r.s. Then $\forall t \geq 1 : \Pr(T \geq t) \leq \frac{E[T]}{t}$

Also, if $\exists b : \Pr(T > b) = 0$, then

$$\text{Histss: } \Pr(T \geq t) \geq \frac{E[T] - t}{b - t}$$

BOUNDED VARIABLES

PROOF

Recall that

$$E[T] = \sum_{x=1}^{\infty} \Pr(T \geq x)$$

Thus

$$E[T] \geq \sum_{x=1}^t \Pr(T \geq x) \geq t \Pr(T \geq t)$$

If $\Pr(T > b) = 0$:

$$\begin{aligned} E[T] &= \sum_{x=1}^b \Pr(T \geq x) + \sum_{x=b+1}^{\infty} \Pr(T \geq x) \\ &\leq t + (b-t) \Pr(T \geq t) \\ \Rightarrow \Pr(T \geq t) &\geq \frac{E[T] - t}{b - t} \end{aligned}$$

EXERCISE Prove that, if X is not the constant r.s. $X = 0$ ($\Pr(X=0) = 1$) then

$$\Pr(X > t) < \frac{E[X]}{t}$$

Suppose that we have 2 high-probability bounds on the tail of $T_A(u)$:

$$\Pr(T_{A^*}(u) \geq C \cdot f(u)) \leq \frac{1}{n^d}$$

Hp:

- Assume that A^* has a deterministic polynomial upper bound on running time:

$$\Pr(T_{A^*}(u) \leq n^{\alpha}) = 1, \text{ with } \alpha \leq d$$

- this is often the case: it is rare that a good randomized algorithm has exponential execution (consider e.g., **QUICKSORT**: $\Pr(T_{QS}(u) \leq c \cdot u^2) = 1$)

- α can usually be chosen as a function of C in $\Pr(T_A(u) \geq C \cdot f(u))$

$$\alpha = \alpha_C \quad (\text{increasing function}).$$

Thus α can be made $\geq \alpha$ (e.g. MR)

From Markov's Lemma (2), letting

$$t = C \cdot f(u) \text{ and } b = n^\alpha :$$

$$E[T] \leq C \cdot f(u) + \frac{(n^\alpha - C \cdot f(u))}{n^\alpha} \leq$$

$$\begin{aligned} E[T] &\leq \\ \Pr(X \geq t)(b-t) &+ \end{aligned}$$

$$\leq C \cdot f(u) + 1 = O(f(u))$$

Thus, englysus w.h. p. also yields the englysisunotuation as a by-product

Markov's Lemma also provides an upper bound on $\Pr(T(u) \geq t)$ as a function of $E[T(u)]$ but it is very weak:

$$\Pr(T(u) \geq k E[T(u)]) \leq \frac{1}{k}$$

To achieve high probability we are forced to set $k = m^d$, thus increasing the bound on $T(u)$ by a factor m^{d+1} .

HOLGER: High probability analysis is generally more powerful than analysis in expectation and is usually preferable.

We will see that there are some special distributions for which we can get more stringent concentration bounds, showing that the probability mass is much more clustered around the expectation than what is implied by Markov's Lemma. Important tool in randomized analysis!

CHEBNOFF'S BOUNDS

CONTEXT: Analysis often relies on indicator variables: Ω_1 random variables X_i : with $\Pr(X_i = 1) = p_i = \mathbb{E}[X_i]$

EXAMPLE: For MAX-3-CNF-SAT

Isism: $y_j^i = 1 \Leftrightarrow$ under random assignment $\vec{y}, y_j^i|_{\vec{y}} = \text{true}$

Multiple indicator variables, when mutually independent are also called Bernoulli trials.

Bernoulli trials are often employed in the analysis of randomized algorithms: the running time can often be rewritten as sum of Bernoulli trials.

Let X_1, X_2, \dots, X_n be mutually independent indicator variables (Bernoulli trials) with $\Pr(X_i = 1) = p_i = \mathbb{E}[X_i]$. We need to study

$$X = \sum_{i=1}^n X_i.$$

We know that

$$\mu = \mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p_i$$

For high-probability analyses we might want to bound $\Pr(X > (1 + \delta)\mu)$

that is, the probability that the sum-variable X deviates from its expectation by more than a $(1+\delta)$ factor (or, analogously, $\Pr(X < (-\delta)\mu), 0 \leq \delta \leq 1$).

Hoeffding's lemma already provides a bound. Since X is nonnegative and nonconstant:

$$\Pr(X > (1+\delta)\mu) \leq \frac{\mu}{(1+\delta)\mu} = \frac{1}{1+\delta}$$

Chebyshev's bounds will provide a much more stringent bound on $\Pr(X > (1+\delta)\mu)$ when X is the sum of Bernoulli trials.
 → very useful for high-probability analyses

CHEBYSHEV'S BOUND : Let X_1, X_2, \dots, X_n be independent indicator variables (Bernoulli-trials) with

$\Pr(X_i = 1) = p_i = E[X_i], 0 < p_i < 1, 1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i, \mu = E[X] = \sum_{i=1}^n p_i$

Then $\forall \delta > 0$:

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

To appreciate the difference with Hoeffding, let $p_i = \frac{1}{2}, 1 \leq i \leq n \Rightarrow \mu = \sum_{i=1}^n \frac{1}{2} = \frac{n}{2}$ - (in this case, X is a binomial variable $X \sim \text{Bin}(n, \frac{1}{2})$)

Let $\delta = \frac{3}{4}$. Then, from Markov's Lemma:

$$\Pr(X > (1 + \frac{3}{4})\mu) = \Pr(X > \frac{7}{8}\mu) < \frac{1}{2}/(\frac{7}{8}\mu) = 4/7$$

From Chernoff's bound 1:

$$\Pr(X > (1 + \frac{3}{4})\frac{\mu}{2}) < \left(\frac{e^{3/4}}{e^{7/4}}\right)^{\frac{\mu}{2}} < \left(\frac{9}{10}\right)^{\frac{\mu}{2}}$$

OBSERVATION: Markov provides a constant upper bound. Chernoff gives an upper bound going to 0 exponentially fast in μ !

(e.g.: $\mu = 100$: $\left(\frac{9}{10}\right)^{\frac{\mu}{2}} < 2.7 \cdot 10^{-5}$)

Important for high-probability analysis.

PROOF We study a derived variable

$Y = f(X)$ trying to obtain a tighter bound.

Let $t > 0$ be a parameter (to be chosen by the analysis to optimise the bound), and set

$$Y = Y_t = e^{tX}$$

(new r.v. obtained by exponentiating the values of X):

$$\begin{aligned} \Pr(X > (1 + \delta)\mu) &= \Pr(tX > t(1 + \delta)\mu) \\ &= \Pr(e^{tX} > e^{t(1 + \delta)\mu}) = \Pr(Y > e^{t(1 + \delta)\mu}) \end{aligned}$$

(the event " $X > (t+\delta)\mu$ " is the same as the event " $Y > e^{t+\delta}\mu$ ")

Let us apply Harkor's Lemma to Y :

$$\Pr(Y > e^{t+\delta}\mu) \leq \frac{E[Y]}{e^{t+\delta}\mu} \quad (1)$$

Let us now study $E[Y]$:

$$\begin{aligned} E[Y] &= E[e^{tX}] = E[e^{t\sum_i X_i}] \\ &= E[\prod_{i=1}^n e^{tX_i}] \end{aligned}$$

Since the X_i are mutually independent, so are the derived variables e^{tX_i} .

We know that the expectation of a product of mutually independent variables is the product of the expectations:

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

Let us now study $E[e^{tX_i}]$:

$$E[e^{tX_i}] = \sum_{x_i=0}^1 e^{t \cdot 0} \cdot \Pr(X_i=0) + e^{t \cdot 1} \cdot \Pr(X_i=1)$$

$$= 1 - p_i + e^t p_i = 1 + p_i(e^t - 1)$$

Therefore:

$$\prod_{i=1}^n E[e^{tX_i}] = \prod_{i=1}^n (1 + p_i(e^t - 1))$$

Recall that

$$1 + z < e^z, \quad \forall z > 0$$

Let us apply the bound to each term of the product, with

$$z_i = p_i(e^t - 1);$$

$$\prod_{i=1}^n E[e^{tX_i}] < \prod_{i=1}^n e^{p_i(e^t - 1)} = e^{\sum_{i=1}^n p_i(e^t - 1)}$$

$$= e^{(e^t - 1) \sum_{i=1}^n p_i} = e^{(e^t - 1)\mu}$$

Plugging in (1) we have shown that

$$\Pr(X > (1+\delta)\mu) < \frac{E[X]}{e^{(1+\delta)\mu}} < \frac{e^{(e^t - 1)\mu}}{e^{E(1+\delta)\mu}}$$

This holds $\forall t > 0$. To obtain the best bound, we find the minimum value of $f(t) =$
 $e^{(e^t - 1)\mu}$ for $t > 0$

$$\frac{d}{dt} f(t) = 0 \Leftrightarrow e^t \cdot \mu - (\nu + \delta) \mu = 0$$

$$\Leftrightarrow t = \ln(\nu + \delta)$$

We obtain

$$\Pr(X > (\nu + \delta)\mu) \leq \frac{e^{(\ln(\nu + \delta)) - s)\mu}}{\frac{\ln(\nu + \delta)(\nu + \delta)\mu}{e}}$$

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

CHECK PASSAGES AT HOME

CHEBYSHEV'S bound 2: Let X_1, X_2, \dots, X_n be independent indicator variables (Bernoulli-trials) with $\Pr(X_i = 1) = p_i = E[X_i]$, $0 < p_i < 1$, $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$, $\mu = E[X] = \sum_{i=1}^n p_i$.

Then $\forall 0 < \delta < 1$:

$$\Pr(X > (\nu + \delta)\mu) < e^{-\delta^2 \mu / 3}$$

PROOF We can rewrite the statement at Chebyshev bound 1 as:

$$\Pr(X > (\nu + \delta)\mu) < e^{[\delta^2 - (\nu + \delta)\ln(\nu + \delta)]\mu}$$

Taylor's series for $\ln(1+\delta)$:

$$\ln(1+\delta) = \sum_{i=1}^{\infty} (-1)^{i+1} \delta^i / i$$

$$\Rightarrow \ln(1+\delta) = \sum_{i=1}^{\infty} (-1)^i \delta^i / i$$

thus

$$(1+\delta)\ln(1+\delta) = \ln(1+\delta) + \delta(\ln(1+\delta)) =$$

$$= S + \sum_{i=2}^{\infty} (-1)^i \delta^i \left(\frac{1}{i-1} - \frac{1}{i} \right) \quad (\text{exercise})$$

decreasing in i

Stopping at $i=3$:

$$(1+\delta)\ln(1+\delta) > S + \frac{\delta^2}{2} - \frac{\delta^3}{6} \geq S + \frac{\delta^2}{2} - \frac{\delta^2}{6} = S + \frac{\delta^2}{3}$$

Therefore

$$\Pr(X > (1+\delta)\mu) < e^{(S - S - S/3)\mu} = e^{-\delta^2\mu/3}$$

Similarly (no proof):

CHEBYSHEV'S SOUND 3: let X_1, X_2, \dots, X_n be independent indicator variables (Bernoulli-trials) with $\Pr(X_i = 1) = p_i = E[X_i]$, $0 < p_i < 1$, $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$, $\mu = E[X] = \sum_{i=1}^n p_i$.

then $\forall 0 < \delta < 1$:

$$\Pr(X < (1-\delta)\mu) < e^{-\delta^2\mu/2}$$

How to use Chebyshev's Sound: if we can make $\delta^2\mu/2 = d \ln n$ we obtain

$$\Pr(X > (1+\delta)\mu) < \frac{1}{n^d}$$