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 <u>worst-case</u> complexity
  - $E[T_A(n)] = \Theta(g(n))$

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:
  $Pr(\forall i : A(i)) = 1$
  These are called LAS VEGAS algorithms
  Randomization only affects running time
  EXAMPLE: RANDOMIZED QUICKSORT
- Algorithms that can be incorrect:
  $Pr(\forall i : A(i)) > 0$
  Randomization affects correctness and may affect running time (or not)
  EXAMPLE: MILLER-RABIN PRIMALITY-TEST
  KARGER'S MIN CUT ALGORITHM
  These are called MONTE CARLO algorithms

- subclasses: 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 defeat worst-case scenarios (e.g. symmetry breaking in the exponential backoff ethernet protocol)
- Simplicity is often counterbalanced by increased complexity in the analysis

RUNNING THE 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_{RS} has running time $T(n) = O(f(n))$ with high probability (w.h.p.) if

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

  $$Pr(T_{A_{RS}}(n) \geq c \cdot f(n)) \leq \frac{1}{n^d}$$

• CORRECTNESS ANALYSIS

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

$$\exists d>0, n_0 \in \mathbb{N} : \forall n \geq n_0 : \text{Miller}(n) : \Pr(i \not\in A_\pi(i)) \leq \frac{1}{n^d}$$

REMARK By choosing $s = d \log_2 |\langle n \rangle|$ MILLER-RABIN can be made correct w.h.p:

$$\Pr(MR(n) \text{ incorrect}) \leq \frac{1}{2^s} = \frac{1}{2^{d \log_2 |\langle n \rangle|}} = \frac{1}{|\langle n \rangle|^d}$$

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

We will use Markov'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.v. Then $\forall t \geq 1 : \Pr(T \geq t) \leq \frac{E[T]}{t}$

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

$\forall 1 \leq t \leq b : \Pr(T \geq t) \geq \frac{E[T] - t}{b - t}$ 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 $\Pr(T > b) = 0$:

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

$$\leq t + (b - t) \Pr(T \geq t)$$

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

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

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

Suppose that we have a high-probability bound on the tail of $T_{A_\pi}(n)$:

$$\Pr(T_{A_\pi}(n) \geq c \cdot f(n)) \leq \frac{1}{n^d}$$

- Assume that A_π has a deterministic polynomial upper bound on running time:

$$\Pr(T_{A_\pi}(n) \leq n^a) = 1, \text{ with } a \leq d.$$

  - this is often the case: it is rare that a good randomized algorithm has expensive executions (consider e.g., QUICKSORT: $\Pr(T_{QS}(n) \leq c \cdot n^2) = 1$)
  - d can usually be chosen as a function of c in $\Pr(T_A(n) \geq c \cdot f(n))$ $d = d_c$ (increasing function). Thus d can be made $\geq a$ (e.g. MR)

From Markov's lemma (2), setting $t = c \cdot f(n)$ and $b = n^d$:

$$E[T] \leq c \cdot f(n) + \frac{(n^d - c \cdot f(n))}{n^d} \leq$$

$E[T] \leq \Pr(X \geq t)(b-t) + t$

$$\leq c \cdot f(n) + 1 = O(f(n))$$

Thus, analysis w.h.p. also yields the analysis in expectation 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 $k = n^d$, thus increasing the bound on $T(u)$ by a factor n^d!

MORAL: 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: CHERNOFF'S BOUNDS

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

EXAMPLE: For MAX-3-CNF-SAT

$1 \le i \le m$, $1 \le j \le 3$: $Y_j^i = 1$ iff under random assignment β, $y_j^i|_\beta = \text{true}$

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 a sum of Bernoulli trials.

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

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

We know that

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

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, $\Pr(X < (1-\delta)\mu)$, $0 < \delta < 1$).

Markov's Lemma already provides a bound. Since X is nonnegative and nonconstant:

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

Chernoff'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

CHERNOFF'S BOUND 1: Let $X_1, X_2, \ldots, X_n$ be independent indicator variables (Bernoulli trials) with $\Pr(X_i = 1) = p_i = E[X_i]$, $0 < p_i < 1$, $1 \le i \le n$. Let $X = \sum_{i=1}^{n} X_i$, $\mu = E[X] = \sum_{i=1}^{n} p_i$

Then $\forall \delta > 0$:

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

To appreciate the difference with Markov, let $p_i = \frac{1}{2}$, $1 \le i \le n \Rightarrow \mu = \sum_{i=1}^{n} \frac{1}{2} = \frac{n}{2}$ - (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:

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

From Chernoff's bound 1:

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

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

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

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 optimize the bound), and set

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

(derives r.v. obtained by exponentiating the value of X):

$$\Pr(X > (1+\delta)\mu) = \Pr\left(tX > t(1+\delta)\mu\right)$$

$$= \Pr\left(e^{tX} > e^{t(1+\delta)\mu}\right) = \Pr\left(Y > e^{t(1+\delta)\mu}\right)$$

(the event "$X \geq (1+\delta)\mu$" is the same as the event "$Y \geq e^{t(1+\delta)\mu}$")

Let us apply Markov's Lemma to Y:

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

Let us now study $E[Y]$:

$$E[Y] = E[e^{tX}] = E[e^{t\sum_{i=1}^{n} X_i}]$$

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

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 $E[e^{tX_i}]$:

$$E[e^{tX_i}] = 1 = \underbrace{e^{t\cdot 0}}_{Pr(X_i=0)} \cdot (1-p_i) + \underbrace{e^{t\cdot 1}}_{Pr(X_i=1)} \cdot p_i$$

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

Therefore:

$$\prod_{i=1}^{n} E[e^{t X_i}] = \prod_{i=1}^{n} \left(1 + \underbrace{p_i (e^t - 1)}_{z_i}\right)$$

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

Let us apply the bound to the terms of the product, with $z_i = p_i (e^t - 1)$:

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

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

Plugging in (1) we have shown that

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

This holds $\forall t > 0$. To obtain the best bound, we find the minimum value of $f(t) =$ for $t > 0$

$$\frac{d}{dt} f(t) = 0 \iff e^t \cdot \mu - (1+\delta)\mu = 0$$

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

We obtain

$$\Pr(X > (1+\delta)\mu) < \frac{e^{(e^{\ln(1+\delta)} - 1)\mu}}{\ln(1+\delta)(1+\delta)\mu} =$$

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

CHECK PASSAGES AT HOME

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

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

$$\text{Prob}(X > (1+\delta)\mu) < e^{-\delta^2 \mu / 3}$$

PROOF We can rewrite the statement of Chernoff bound 1 as:

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

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

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

Thus

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

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

Stopping at $i=3$:

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

Therefore

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

Similarly (no proof):

CHERNOFF'S BOUND 3: Let $X_1, X_2, \ldots, X_n$ 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 $\forall\, 0 < \delta < 1$:

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

How to use Chernoff's bounds: if we can make $\delta^2\mu/3 = d \ln n$ we obtain

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