Introduction to Randomized Algorithms: QuickSort

Lecture 2

January 17, 2019

Outline

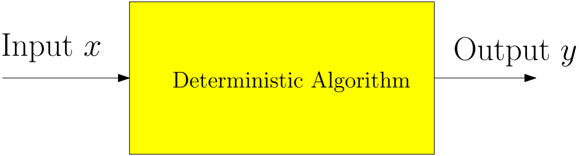
Our goal

- Basics of randomization – probability space, expectation, events, random variables, etc.
- Randomized Algorithms – Two types
  - Las Vegas
  - Monte Carlo
- Randomized Quick Sort

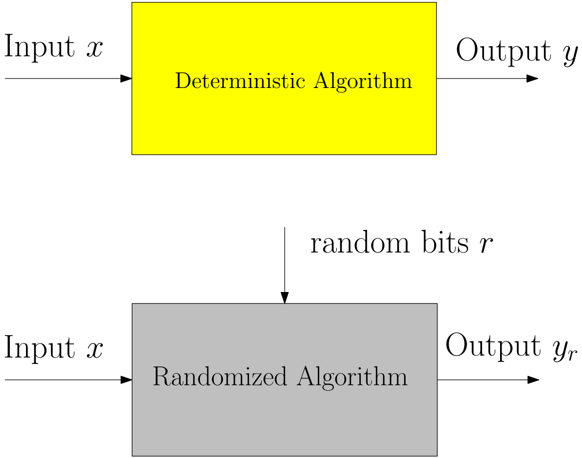
Part I

Introduction to Randomized Algorithms

Randomized Algorithms



Randomized Algorithms



Example: Randomized **QuickSort**

QuickSort ?

1. Pick a pivot element from array
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Randomized **QuickSort**

1. Pick a pivot element **uniformly at random** from the array
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Example: Randomized Quicksort

Recall: **QuickSort** can take $\Omega(n^2)$ time to sort array of size n.

Example: Randomized Quicksort

Recall: **QuickSort** can take $\Omega(n^2)$ time to sort array of size n.

Theorem

Randomized **QuickSort** *sorts a given array of length* n *in* $O(n \log n)$ *expected time.*

Example: Randomized Quicksort

Recall: **QuickSort** can take $\Omega(n^2)$ time to sort array of size n.

Theorem

Randomized **QuickSort** *sorts a given array of length* n *in* $O(n \log n)$ *expected time.*

Note: On *every* input randomized **QuickSort** takes $O(n \log n)$ time in expectation. On *every* input it may take $\Omega(n^2)$ time with some small probability.

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Deterministic algorithm:

1. Multiply A and B and check if equal to C.
2. Running time?

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Deterministic algorithm:

1. Multiply A and B and check if equal to C.
2. Running time? $O(n^3)$ by straight forward approach. $O(n^{2.37})$ with fast matrix multiplication (complicated and impractical).

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Randomized algorithm:

1. Pick a random $n \times 1$ vector r.
2. Return the answer of the equality $ABr = Cr$.
3. Running time?

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Randomized algorithm:

1. Pick a random $n \times 1$ vector r.
2. Return the answer of the equality $ABr = Cr$.
3. Running time? $O(n^2)$!

Example: Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices A, B, C is $AB = C$?

Randomized algorithm:

1. Pick a random $n \times 1$ vector r.
2. Return the answer of the equality $ABr = Cr$.
3. Running time? $O(n^2)$!

Theorem

If $AB = C$ then the algorithm will always say YES. If $AB \neq C$ then the algorithm will say YES with probability at most $1/2$. Can repeat the algorithm 100 times independently to reduce the probability of a false positive to $1/2^{100}$.

Why randomized algorithms?

1. Many many applications in algorithms, data structures and computer science!
2. In some cases only known algorithms are randomized or randomness is provably necessary.
3. Often randomized algorithms are (much) simpler and/or more efficient.
4. Several deep connections to mathematics, physics etc.
5. ...
6. Lots of fun!

Average case analysis vs Randomized algorithms

Average case analysis:

1. Fix a deterministic algorithm.
2. Assume inputs comes from a probability distribution.
3. Analyze the algorithm's *average* performance over the distribution over inputs.

Randomized algorithms:

1. Algorithm uses random bits in addition to input.
2. Analyze algorithms *average* performance over the given input where the average is over the random bits that the algorithm uses.
3. On each input behaviour of algorithm is random. Analyze worst-case over all inputs of the (average) performance.

Part II

Basics of Discrete Probability

Discrete Probability

We restrict attention to finite probability spaces.

Definition

A discrete probability space is a pair $(\mathbf{\Omega}, \mathbf{Pr})$ consists of finite set $\mathbf{\Omega}$ of **elementary events** and function $p : \mathbf{\Omega} \rightarrow [\mathbf{0}, \mathbf{1}]$ which assigns a probability $\mathbf{Pr}[\omega]$ for each $\omega \in \mathbf{\Omega}$ such that $\sum_{\omega \in \mathbf{\Omega}} \mathbf{Pr}[\omega] = \mathbf{1}$.

Discrete Probability

We restrict attention to finite probability spaces.

Definition

A discrete probability space is a pair $(\Omega, \Pr)$ consists of finite set Ω of **elementary events** and function $p : \Omega \to [0, 1]$ which assigns a probability $\Pr[\omega]$ for each $\omega \in \Omega$ such that $\sum_{\omega \in \Omega} \Pr[\omega] = 1$.

Example

An unbiased coin. $\Omega = \{H, T\}$ and $\Pr[H] = \Pr[T] = 1/2$.

Example

A 6-sided unbiased die. $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\Pr[i] = 1/6$ for $1 \leq i \leq 6$.

Events

Definition

Given a probability space $(\Omega, \Pr)$ an **event** is a subset of Ω. In other words an event is a collection of elementary events. The probability of an event A, denoted by $\Pr[A]$, is $\sum_{\omega \in A} \Pr[\omega]$.

The **complement event** of an event $A \subseteq \Omega$ is the event $\Omega \setminus A$ frequently denoted by $\bar{A}$.

Events

Definition

Given a probability space $(\Omega, \Pr)$ an **event** is a subset of Ω. In other words an event is a collection of elementary events. The probability of an event A, denoted by $\Pr[A]$, is $\sum_{\omega \in A} \Pr[\omega]$.

The **complement event** of an event $A \subseteq \Omega$ is the event $\Omega \setminus A$ frequently denoted by $\bar{A}$.

Example

A pair of independent dice. $\Omega = \{(i,j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$.
Let A be the event that the sum of the two numbers on the dice is even.
Then $A = \left\{ (i,j) \in \Omega \;\middle|\; (i+j) \text{ is even} \right\}$.
$\Pr[A] = |A|/36 = 1/2$.

Independent Events

Definition

Given a probability space $(\Omega, \mathbf{Pr})$ and two events A, B are **independent** if and only if $\mathbf{Pr}[A \cap B] = \mathbf{Pr}[A]\,\mathbf{Pr}[B]$. Otherwise they are *dependent*. In other words A, B independent implies one does not affect the other.

Independent Events

Definition

Given a probability space $(\Omega, \Pr)$ and two events A, B are **independent** if and only if $\Pr[A \cap B] = \Pr[A] \Pr[B]$. Otherwise they are *dependent*. In other words A, B independent implies one does not affect the other.

Example

Two coins. $\Omega = \{HH, TT, HT, TH\}$ and $\Pr[HH] = \Pr[TT] = \Pr[HT] = \Pr[TH] = 1/4$.

1. A is the event that the first coin is heads and B is the event that second coin is tails. A, B are independent.

Independent Events

Definition

Given a probability space $(\Omega, \Pr)$ and two events A, B are **independent** if and only if $\Pr[A \cap B] = \Pr[A]\Pr[B]$. Otherwise they are *dependent*. In other words A, B independent implies one does not affect the other.

Example

Two coins. $\Omega = \{HH, TT, HT, TH\}$ and $\Pr[HH] = \Pr[TT] = \Pr[HT] = \Pr[TH] = 1/4$.

1. A is the event that the first coin is heads and B is the event that second coin is tails. A, B are independent.
2. A is the event that both are not tails and B is event that second coin is heads.

Independent Events

Definition

Given a probability space $(\Omega, \Pr)$ and two events A, B are **independent** if and only if $\Pr[A \cap B] = \Pr[A]\Pr[B]$. Otherwise they are *dependent*. In other words A, B independent implies one does not affect the other.

Example

Two coins. $\Omega = \{HH, TT, HT, TH\}$ and $\Pr[HH] = \Pr[TT] = \Pr[HT] = \Pr[TH] = 1/4$.

1. A is the event that the first coin is heads and B is the event that second coin is tails. A, B are independent.
2. A is the event that both are not tails and B is event that second coin is heads. A, B **are dependent**.

Union bound

The probability of the union of two events, is no bigger than the probability of the sum of their probabilities.

Lemma

For any two events $\mathcal{E}$ and $\mathcal{F}$, we have that
$\mathbf{Pr}\left[\mathcal{E} \cup \mathcal{F}\right] \leq \mathbf{Pr}\left[\mathcal{E}\right] + \mathbf{Pr}\left[\mathcal{F}\right]$.

Proof.

Consider $\mathcal{E}$ and $\mathcal{F}$ to be a collection of elmentery events (which they are). We have

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{E} \cup \mathcal{F}\right] &= \sum_{x \in \mathcal{E} \cup \mathcal{F}} \mathbf{Pr}[x] \\ &\leq \sum_{x \in \mathcal{E}} \mathbf{Pr}[x] + \sum_{x \in \mathcal{F}} \mathbf{Pr}[x] = \mathbf{Pr}\left[\mathcal{E}\right] + \mathbf{Pr}\left[\mathcal{F}\right].\end{aligned}$$

Random Variables

Definition

Given a probability space $(\Omega, \Pr)$ a (real-valued) random variable X over Ω is a function that maps each elementary event to a real number. In other words $X : \Omega \to \mathbb{R}$.

Random Variables

Definition

Given a probability space $(\Omega, \Pr)$ a (real-valued) random variable X over Ω is a function that maps each elementary event to a real number. In other words $X : \Omega \to \mathbb{R}$.

Definition

Expectation For a random variable X over a probability space $(\Omega, \Pr)$ the **expectation** of X is defined as $\sum_{\omega \in \Omega} \Pr[\omega] X(\omega)$. In other words, the expectation is the average value of X according to the probabilities given by $\Pr[\cdot]$.

Expectation

Example

A 6-sided unbiased die. $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\Pr[i] = 1/6$ for $1 \leq i \leq 6$.

1. $X : \Omega \to \mathbb{R}$ where $X(i) = i \mod 2$. Then $E[X] = \sum_{i=1}^{6} \Pr[i] \cdot X(i) = \frac{1}{6} \sum_{i=1}^{6} X(i) = 1/2$.

Expectation

Example

A 6-sided unbiased die. $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\Pr[i] = 1/6$ for $1 \leq i \leq 6$.

1. $X : \Omega \to \mathbb{R}$ where $X(i) = i \mod 2$. Then $\mathbf{E}[X] = \sum_{i=1}^{6} \Pr[i] \cdot X(i) = \frac{1}{6} \sum_{i=1}^{6} X(i) = 1/2$.
2. $Y : \Omega \to \mathbb{R}$ where $Y(i) = i^2$. Then $\mathbf{E}[Y] = \sum_{i=1}^{6} \frac{1}{6} \cdot i^2 = 91/6$.

Expected number of vertices?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. Compute the expected number of vertices in H.

(A) n/2.

(B) n/4.

(C) m/2.

(D) m/4.

(E) none of the above.

Expected number of vertices is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\mathbf{Pr}[\omega] = \frac{1}{2^n}$.

Expected number of vertices is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\Pr[\omega] = \frac{1}{2^n}$.
- $X(\omega) = \#$ vertices in H as per $\omega = \#$ 1s in ω.

Expected number of vertices is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\Pr[\omega] = \frac{1}{2^n}$.
- $X(\omega) = \#$ vertices in H as per $\omega = \#$ 1s in ω.

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{\omega \in \Omega} \Pr[\omega]\, X(\omega) \\
&= \sum_{\omega \in \Omega} {}^{1}/_{2^n} X(\omega) \\
&= {}^{1}/_{2^n} \sum_{k=0}^{n} \binom{n}{k} k \\
&= {}^{1}/_{2^n} (2^n \tfrac{n}{2}) \\
&= n/2
\end{aligned}
$$

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

(A) n/2.

(B) n/4.

(C) m/2.

(D) m/4.

(E) none of the above.

Expected number of edges is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\Pr[\omega] = \frac{1}{2^n}$.

Expected number of edges is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\Pr[\omega] = \frac{1}{2^n}$.
- $X(\omega) = \#$ edges present in H as per ω = **??**

Expected number of edges is:

Probability Space

- $\Omega = \{0,1\}^n$. For $\omega \in \{0,1\}^n$, $\omega_v = 1$ if vertex v is present in H, else is zero.
- For each $\omega \in \Omega$, $\Pr[\omega] = \frac{1}{2^n}$.
- $X(\omega) = \#$ edges present in H as per ω = ??

How to compute $\mathbf{E}[X]$?

Indicator Random Variables

Definition

A **binary random variable** is one that takes on values in $\{0, 1\}$.

Indicator Random Variables

Definition

A **binary random variable** is one that takes on values in $\{0, 1\}$.

Special type of random variables that are quite useful.

Definition

Given a probability space $(\Omega, \Pr)$ and an event $A \subseteq \Omega$ the indicator random variable X_A is a binary random variable where $X_A(\omega) = 1$ if $\omega \in A$ and $X_A(\omega) = 0$ if $\omega \notin A$.

Indicator Random Variables

Definition

A **binary random variable** is one that takes on values in $\{0, 1\}$.

Special type of random variables that are quite useful.

Definition

Given a probability space $(\Omega, \Pr)$ and an event $A \subseteq \Omega$ the indicator random variable X_A is a binary random variable where $X_A(\omega) = 1$ if $\omega \in A$ and $X_A(\omega) = 0$ if $\omega \not\in A$.

Example

A 6-sided unbiased die. $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\Pr[i] = 1/6$ for $1 \leq i \leq 6$. Let A be the even that i is divisible by 3. Then $X_A(i) = 1$ if $i = 3, 6$ and 0 otherwise.

Expectation of indicator random variable

Proposition

For an indicator variable X_A, $\mathbf{E}[X_A] = \Pr[A]$.

Proof.

$$
\begin{aligned}
\mathbf{E}[X_A] &= \sum_{\omega \in \Omega} X_A(\omega) \Pr[\omega] \\
&= \sum_{\omega \in A} 1 \cdot \Pr[\omega] + \sum_{\omega \in \Omega \setminus A} 0 \cdot \Pr[\omega] \\
&= \sum_{\omega \in A} \Pr[\omega] \\
&= \Pr[A] .
\end{aligned}
$$

$\square$

Linearity of Expectation

Lemma

Let X, Y *be two random variables (not necessarily independent) over a probability space* $(\Omega, \Pr)$. *Then* $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.

Proof.

$$\begin{aligned}\mathbf{E}[X+Y] &= \sum_{\omega\in\Omega} \Pr[\omega]\,(X(\omega)+Y(\omega)) \\ &= \sum_{\omega\in\Omega} \Pr[\omega]\,X(\omega) + \sum_{\omega\in\Omega} \Pr[\omega]\,Y(\omega) = \mathbf{E}[X] + \mathbf{E}[Y].\end{aligned}$$

$\square$

Linearity of Expectation

Lemma

Let X, Y be two random variables (not necessarily independent) over a probability space $(\Omega, \Pr)$. Then $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.

Proof.

$$\begin{aligned}\mathbf{E}[X+Y] &= \sum_{\omega\in\Omega} \Pr[\omega]\,(X(\omega)+Y(\omega)) \\ &= \sum_{\omega\in\Omega} \Pr[\omega]\,X(\omega) + \sum_{\omega\in\Omega} \Pr[\omega]\,Y(\omega) = \mathbf{E}[X] + \mathbf{E}[Y].\end{aligned}$$

□

Corollary

$\mathbf{E}[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = \sum_{i=1}^n a_i\,\mathbf{E}[X_i]$.

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

- Event $A_e =$ edge $e \in E$ is present in H.
- $\mathbf{Pr}\left[A_{e=(u,v)}\right] = \mathbf{Pr}[u \text{ and } v \text{ both are present}] = \mathbf{Pr}[u \text{ is present}] \cdot \mathbf{Pr}[v \text{ is present}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

- Event $A_e =$ edge $e \in E$ is present in H.
- $\mathbf{Pr}\left[A_{e=(u,v)}\right] = \mathbf{Pr}[u \text{ and } v \text{ both are present}] = \mathbf{Pr}[u \text{ is present}] \cdot \mathbf{Pr}[v \text{ is present}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
- X_{A_e} indicator random variables, then $\mathbf{E}[X_{A_e}] = \mathbf{Pr}[A_e]$.

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

- Event A_e = edge $e \in E$ is present in H.
- $\Pr\left[A_{e=(u,v)}\right] = \Pr[u \text{ and } v \text{ both are present}] = \Pr[u \text{ is present}] \cdot \Pr[v \text{ is present}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
- X_{A_e} indicator random variables, then $\mathbf{E}[X_{A_e}] = \Pr[A_e]$.
- Let $X = \sum_{e \in E} X_{A_e}$ (Number of edges in H)

Expected number of edges?

Let $G = (V, E)$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

- Event A_e = edge $e \in E$ is present in H.
- $\Pr\left[A_{e=(u,v)}\right] = \Pr[u \text{ and } v \text{ both are present}] = \Pr[u \text{ is present}] \cdot \Pr[v \text{ is present}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
- X_{A_e} indicator random variables, then $\mathbf{E}[X_{A_e}] = \Pr[A_e]$.
- Let $X = \sum_{e \in E} X_{A_e}$ (Number of edges in H)

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{e \in E} X_{A_e}\right] = \sum_{e \in E} \mathbf{E}[X_{A_e}] = \sum_{e \in E} \Pr[A_e] = \frac{m}{4}$$

Expected number of edges?

Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph with n vertices and m edges. Let H be the graph resulting from independently deleting every vertex of G with probability $1/2$. The expected number of edges in H is

- Event A_e = edge $e \in \mathsf{E}$ is present in H.
- $\mathbf{Pr}\left[A_{e=(u,v)}\right] = \mathbf{Pr}[u \text{ and } v \text{ both are present}] = \mathbf{Pr}[u \text{ is present}] \cdot \mathbf{Pr}[v \text{ is present}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
- X_{A_e} indicator random variables, then $\mathbf{E}[X_{A_e}] = \mathbf{Pr}[A_e]$.
- Let $X = \sum_{e \in \mathsf{E}} X_{A_e}$ (Number of edges in H)

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{e \in \mathsf{E}} X_{A_e}\right] = \sum_{e \in \mathsf{E}} \mathbf{E}[X_{A_e}] = \sum_{e \in \mathsf{E}} \mathbf{Pr}[A_e] = \frac{m}{4}$$

It is important to setup random variables carefully.

Expected number of triangles?

Let $G = (V, E)$ be a graph with n vertices and m edges. Assume G has t triangles (i.e., a triangle is a simple cycle with three vertices). Let H be the graph resulting from deleting independently each vertex of G with probability $1/2$. The expected number of triangles in H is

(A) $t/2$.

(B) $t/4$.

(C) $t/8$.

(D) $t/16$.

(E) none of the above.

Independent Random Variables

Definition

Random variables X, Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \quad \mathbf{Pr}[X = x \wedge Y = y] = \mathbf{Pr}[X = x]\,\mathbf{Pr}[Y = y]$$

.

Independent Random Variables

Definition

Random variables X, Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \quad \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

.

Examples

Two independent un-biased coin flips: $\Omega = \{HH, HT, TH, TT\}$.

- $X = 1$ if first coin is H else 0. $Y = 1$ if second coin is H else 0.

Independent Random Variables

Definition

Random variables X, Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \quad \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

.

Examples

Two independent un-biased coin flips: $\Omega = \{HH, HT, TH, TT\}$.

- $X = 1$ if first coin is H else 0. $Y = 1$ if second coin is H else 0. Independent.

Independent Random Variables

Definition

Random variables X, Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \quad \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

.

Examples

Two independent un-biased coin flips: $\Omega = \{HH, HT, TH, TT\}$.

- $X = 1$ if first coin is H else 0. $Y = 1$ if second coin is H else 0. Independent.
- $X = \#H$, $Y = \#T$.

Independent Random Variables

Definition

Random variables X, Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \quad \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

.

Examples

Two independent un-biased coin flips: $\Omega = \{HH, HT, TH, TT\}$.

- $X = 1$ if first coin is H else 0. $Y = 1$ if second coin is H else 0. Independent.
- $X = \#H$, $Y = \#T$. Dependent. Why?

Independent Randomized Variables

Lemma

If X and Y are independent then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$

Proof.

$$
\begin{aligned}
\mathbf{E}[X \cdot Y] &= \sum_{\omega \in \Omega} \Pr[\omega] \left(X(\omega) \cdot Y(\omega)\right) \\
&= \sum_{x,y \in \mathbb{R}} \Pr[X = x \wedge Y = y] \left(x \cdot y\right) \\
&= \sum_{x,y \in \mathbb{R}} \Pr[X = x] \cdot \Pr[Y = y] \cdot x \cdot y \\
&= \Big(\sum_{x \in \mathbb{R}} \Pr[X = x]\, x\Big)\Big(\sum_{y \in \mathbb{R}} \Pr[Y = y]\, y\Big) = \mathbf{E}[X]\,\mathbf{E}[Y]
\end{aligned}
$$

□

Types of Randomized Algorithms

Typically one encounters the following types:

1. **Las Vegas randomized algorithms:** for a given input x output of *algorithm is always correct* but the *running time is a random variable*. In this case we are interested in analyzing the *expected* running time.

Types of Randomized Algorithms

Typically one encounters the following types:

1. **Las Vegas randomized algorithms:** for a given input x output of *algorithm is always correct* but the *running time is a random variable*. In this case we are interested in analyzing the *expected* running time.
2. **Monte Carlo randomized algorithms:** for a given input x the *running time is deterministic* but the *output is random*; correct with some probability. In this case we are interested in analyzing the *probability* of the correct output (and also the running time).
3. Algorithms whose running time and output may both be random.

Analyzing Las Vegas Algorithms

Deterministic algorithm Q for a problem Π:

1. Let $Q(x)$ be the time for Q to run on input x of length $|x|$.
2. Worst-case analysis: run time on worst input for a given size n.

$$T_{wc}(n) = \max_{x:|x|=n} Q(x).$$

Analyzing Las Vegas Algorithms

Deterministic algorithm Q for a problem Π:

1. Let $Q(x)$ be the time for Q to run on input x of length $|x|$.
2. Worst-case analysis: run time on worst input for a given size n.

$$T_{wc}(n) = \max_{x:|x|=n} Q(x).$$

Randomized algorithm R for a problem Π:

1. Let $R(x)$ be the time for Q to run on input x of length $|x|$.
2. $R(x)$ is a random variable: depends on random bits used by R.
3. $\mathbf{E}[R(x)]$ is the expected running time for R on x
4. Worst-case analysis: expected time on worst input of size n

$$T_{rand-wc}(n) = \max_{x:|x|=n} \mathbf{E}[R(x)].$$

Analyzing Monte Carlo Algorithms

Randomized algorithm M for a problem Π:

1. Let $M(x)$ be the time for M to run on input x of length $|x|$. For Monte Carlo, assumption is that run time is deterministic.
2. Let $\Pr[x]$ be the probability that M is correct on x.
3. $\Pr[x]$ is a random variable: depends on random bits used by M.
4. Worst-case analysis: success probability on worst input

$$P_{rand-wc}(n) = \min_{x:|x|=n} \Pr[x].$$

Part III

Randomized Quick Sort

Randomized **QuickSort**

Randomized **QuickSort**

1. Pick a pivot element *uniformly at random* from the array.
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Analysis

What events to count?

- Number of Comparisions.

Analysis

What events to count?

- Number of Comparisions.

What is the probability space?

- All the coin tosses at all levels and parts of recursion.

Analysis

What events to count?

- Number of Comparisions.

What is the probability space?

- All the coin tosses at all levels and parts of recursion.

Too Big!!

Analysis

What events to count?

- Number of Comparisions.

What is the probability space?

- All the coin tosses at all levels and parts of recursion.

Too Big!!

What random variables to define?
What are the events of the algorithm?

Analysis via Recurrence

1. Given array A of size n, let $Q(A)$ be number of comparisons of randomized **QuickSort** on A.
2. Note that $Q(A)$ is a random variable.
3. Let A^i_{left} and A^i_{right} be the left and right arrays obtained if:

Let X_i be indicator random variable, which is set to 1 if pivot is of rank i in A, else zero.

$$Q(A) = n + \sum_{i=1}^{n} X_i \cdot \left(Q(A^i_{\text{left}}) + Q(A^i_{\text{right}}) \right).$$

Analysis via Recurrence

1. Given array A of size n, let $Q(A)$ be number of comparisons of randomized **QuickSort** on A.
2. Note that $Q(A)$ is a random variable.
3. Let A^i_{left} and A^i_{right} be the left and right arrays obtained if:

Let X_i be indicator random variable, which is set to 1 if pivot is of rank i in A, else zero.

$$Q(A) = n + \sum_{i=1}^{n} X_i \cdot \left(Q(A^i_{\text{left}}) + Q(A^i_{\text{right}}) \right).$$

Since each element of A has probability exactly of $1/n$ of being chosen:

$$\mathbf{E}[X_i] = \mathbf{Pr}[\text{pivot has rank } i] = 1/n.$$

Independence of Random Variables

Lemma

Random variables X_i *is independent of random variables* $Q(A^i_{left})$ *as well as* $Q(A^i_{right})$, i.e.

$$\mathbf{E}\left[X_i \cdot Q(A^i_{left})\right] = \mathbf{E}[X_i]\,\mathbf{E}\left[Q(A^i_{left})\right]$$
$$\mathbf{E}\left[X_i \cdot Q(A^i_{right})\right] = \mathbf{E}[X_i]\,\mathbf{E}\left[Q(A^i_{right})\right]$$

Proof.

This is because the algorithm, while recursing on $Q(A^i_{\text{left}})$ and $Q(A^i_{\text{right}})$ uses new random coin tosses that are independent of the coin tosses used to decide the first pivot. Only the latter decides value of X_i. □

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.

We have, for any A:

$$Q(A) = n + \sum_{i=1}^{n} X_i \left(Q(A^i_{\text{left}}) + Q(A^i_{\text{right}}) \right)$$

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.

We have, for any A:

$$Q(A) = n + \sum_{i=1}^{n} X_i \left(Q(A^i_{\text{left}}) + Q(A^i_{\text{right}}) \right)$$

By linearity of expectation, and independence random variables:

$$\mathbf{E}\Big[Q(A)\Big] = n + \sum_{i=1}^{n} \mathbf{E}[X_i] \Big(\mathbf{E}\Big[Q(A^i_{\text{left}})\Big] + \mathbf{E}\Big[Q(A^i_{\text{right}})\Big] \Big).$$

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.

We have, for any A:

$$Q(A) = n + \sum_{i=1}^{n} X_i \left(Q(A^i_{\text{left}}) + Q(A^i_{\text{right}}) \right)$$

By linearity of expectation, and independence random variables:

$$\mathbf{E}\Big[Q(A)\Big] = n + \sum_{i=1}^{n} \mathbf{E}[X_i]\Big(\mathbf{E}\Big[Q(A^i_{\text{left}})\Big] + \mathbf{E}\Big[Q(A^i_{\text{right}})\Big]\Big).$$

$$\Rightarrow \quad \mathbf{E}\Big[Q(A)\Big] \leq n + \sum_{i=1}^{n} \frac{1}{n} \left(T(i-1) + T(n-i) \right).$$

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.

Analysis via Recurrence

Let $T(n) = \max_{A:|A|=n} \mathbf{E}[Q(A)]$ be the worst-case expected running time of randomized **QuickSort** on arrays of size n.
We derived:

$$\mathbf{E}\Big[Q(A)\Big] \leq n + \sum_{i=1}^{n} \frac{1}{n} \left(T(i-1) + T(n-i) \right).$$

Note that above holds for any A of size n. Therefore

$$\max_{A:|A|=n} \mathbf{E}[Q(A)] = T(n) \leq n + \sum_{i=1}^{n} \frac{1}{n} \left(T(i-1) + T(n-i) \right).$$

Solving the Recurrence

$$T(n) \leq n + \sum_{i=1}^{n} \frac{1}{n} \left(T(i-1) + T(n-i) \right)$$

with base case $T(1) = 0$.

Solving the Recurrence

$$T(n) \leq n + \sum_{i=1}^{n} \frac{1}{n}\left(T(i-1) + T(n-i)\right)$$

with base case $T(1) = 0$.

Lemma

$T(n) = O(n \log n)$.

Solving the Recurrence

$$T(n) \leq n + \sum_{i=1}^{n} \frac{1}{n}\left(T(i-1) + T(n-i)\right)$$

with base case $T(1) = 0$.

Lemma

$T(n) = O(n \log n)$.

Proof.

(Guess and) Verify by induction. □

Part IV

Slick analysis of QuickSort

A Slick Analysis of **QuickSort**

Let $Q(A)$ be number of comparisons done on input array A:

1. For $1 \leq i < j < n$ let R_{ij} be the event that rank i element is compared with rank j element.
2. X_{ij} is the indicator random variable for R_{ij}. That is, $X_{ij} = 1$ if rank i is compared with rank j element, otherwise 0.

A Slick Analysis of **QuickSort**

Let $Q(A)$ be number of comparisons done on input array A:

1. For $1 \leq i < j < n$ let R_{ij} be the event that rank i element is compared with rank j element.
2. X_{ij} is the indicator random variable for R_{ij}. That is, $X_{ij} = 1$ if rank i is compared with rank j element, otherwise 0.

$$Q(A) = \sum_{1 \leq i < j \leq n} X_{ij}$$

and hence by linearity of expectation,

$$\mathbf{E}\Big[Q(A)\Big] = \sum_{1 \leq i < j \leq n} \mathbf{E}\Big[X_{ij}\Big] = \sum_{1 \leq i < j \leq n} \mathbf{Pr}\Big[R_{ij}\Big].$$

A Slick Analysis of **QuickSort**

$\boxed{R_{ij} = \text{rank } i \text{ element is compared with rank } j \text{ element.}}$

Question: What is $\mathbf{Pr}[R_{ij}]$?

A Slick Analysis of **QuickSort**

$\boxed{R_{ij} = \text{rank } i \text{ element is compared with rank } j \text{ element.}}$

Question: What is $\mathbf{Pr}[R_{ij}]$?

	7	5	9	1	3	4	8	6
With ranks:	6	4	8	1	2	3	7	5

A Slick Analysis of **QuickSort**

$\boxed{R_{ij} = \text{rank } i \text{ element is compared with rank } j \text{ element.}}$

Question: What is $\Pr[R_{ij}]$?

	7	5	9	1	3	4	8	6
With ranks:	6	4	8	1	2	3	7	5

As such, probability of comparing **5** to **8** is $\Pr[R_{4,7}]$.

A Slick Analysis of **QuickSort**

$\boxed{R_{ij} = \text{rank } i \text{ element is compared with rank } j \text{ element.}}$

Question: What is $\Pr[R_{ij}]$?

	7	5	9	1	3	4	8	6
With ranks:	6	4	8	1	2	3	7	5

1. If pivot too small (say **3** [rank 2]). Partition and call recursively:

   $\boxed{7 \mid 5 \mid 9 \mid 1 \mid 3 \mid 4 \mid 8 \mid 6} \Longrightarrow \boxed{1 \quad 3 \quad 7 \mid 5 \mid 9 \mid 4 \mid 8 \mid 6}$

   Decision if to compare **5** to **8** is moved to subproblem.

A Slick Analysis of **QuickSort**

$\boxed{R_{ij} = \text{rank } i \text{ element is compared with rank } j \text{ element.}}$

Question: What is $\Pr[R_{ij}]$?

	7	5	9	1	3	4	8	6
With ranks:	6	4	8	1	2	3	7	5

1. If pivot too small (say **3** [rank 2]). Partition and call recursively:

7	5	9	1	3	4	8	6

$\Longrightarrow$

1	3	7	5	9	4	8	6

Decision if to compare **5** to **8** is moved to subproblem.

2. If pivot too large (say **9** [rank 8]):

7	5	9	1	3	4	8	6

$\Longrightarrow$

7	5	1	3	4	8	6	9

Decision if to compare **5** to **8** moved to subproblem.

7	5	9	1	3	4	8	6
6	4	8	1	2	3	7	5

As such, probability of comparing **5** to **8** is $\mathbf{Pr}[R_{4,7}]$.

1. If pivot is **5** (rank 4). Bingo!

7	5	9	1	3	4	8	6

$\Longrightarrow$

1	3	4		5		7	9	8	6

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr}[\mathbf{R_{i,j}}]$?

7	5	9	1	3	4	8	6
6	4	8	1	2	3	7	5

As such, probability of comparing **5** to **8** is $\mathbf{Pr}[R_{4,7}]$.

1. If pivot is **5** (rank 4). Bingo!

7	5	9	1	3	4	8	6

$\Longrightarrow$

1	3	4	5	7	9	8	6

2. If pivot is **8** (rank 7). Bingo!

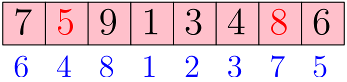
7	5	9	1	3	4	8	6

$\Longrightarrow$

7	5	1	3	4	6	8	9

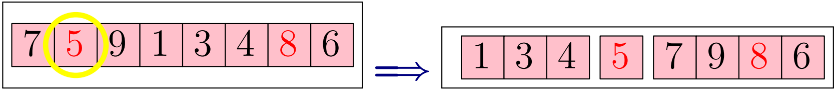
A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr[R_{i,j}]}$?

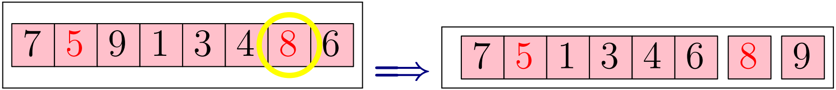


As such, probability of comparing **5** to **8** is $\mathbf{Pr[}R_{4,7}\mathbf{]}$.

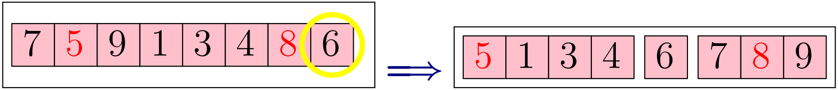
1. If pivot is **5** (rank 4). Bingo!



2. If pivot is **8** (rank 7). Bingo!



3. If pivot in between the two numbers (say **6** [rank 5]):



5 and **8** will never be compared to each other.

A Slick Analysis of **QuickSort**

Question: What is $\textbf{Pr}[\mathbf{R_{i,j}}]$?

Conclusion:

$R_{i,j}$ happens if and only if:

ith or jth ranked element is the first pivot out of ith to jth ranked elements.

How to analyze this?

Thinking acrobatics!

1. Assign every element in the array a random priority (say in $[0, 1]$).
2. Choose pivot to be the element with lowest priority in subproblem.
3. Equivalent to picking pivot uniformly at random (as **QuickSort** do).

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr[R_{i,j}]}$?

How to analyze this?

Thinking acrobatics!

1. Assign every element in the array a random priority (say in $[0, 1]$).
2. Choose pivot to be the element with lowest priority in subproblem.

$\implies$ $R_{i,j}$ happens if either i or j have lowest priority out of elements rank i to j,

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr[R_{i,j}]}$?

How to analyze this?

Thinking acrobatics!

1. Assign every element in the array a random priority (say in $[0, 1]$).
2. Choose pivot to be the element with lowest priority in subproblem.

$\implies$ $R_{i,j}$ happens if either i or j have lowest priority out of elements rank i to j,
There are $k = j - i + 1$ relevant elements.

$$\mathbf{Pr}\left[R_{i,j}\right] = \frac{2}{k} = \frac{2}{j - i + 1}.$$

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr}[R_{ij}]$?

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr}[R_{ij}]$?

Lemma

$\mathbf{Pr}\left[R_{ij}\right] = \frac{2}{j-i+1}$.

A Slick Analysis of **QuickSort**

Question: What is $\mathbf{Pr}[R_{ij}]$?

Lemma

$\mathbf{Pr}\left[R_{ij}\right] = \frac{2}{j-i+1}$.

Proof.

Let $a_1, \ldots, a_i, \ldots, a_j, \ldots, a_n$ be elements of A in sorted order.
Let $S = \{a_i, a_{i+1}, \ldots, a_j\}$
Observation: If pivot is chosen outside S then all of S either in left array or right array.
Observation: a_i and a_j separated when a pivot is chosen from S for the first time. Once separated no comparison.
Observation: a_i is compared with a_j if and only if either a_i or a_j is chosen as a pivot from S at separation... $\square$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\mathbf{Pr}\left[R_{ij}\right] = \frac{2}{j-i+1}$.

Proof.

Let $a_1, \ldots, a_i, \ldots, a_j, \ldots, a_n$ be sort of A. Let $S = \{a_i, a_{i+1}, \ldots, a_j\}$

Observation: a_i is compared with a_j if and only if either a_i or a_j is chosen as a pivot from S at separation.

Observation: Given that pivot is chosen from S the probability that it is a_i or a_j is exactly $2/|S| = 2/(j-i+1)$ since the pivot is chosen uniformly at random from the array. □

How much is this?

$H_n = \sum_{i=1}^{n} \frac{1}{i}$ is the n'th harmonic number

(A) $H_n = \Theta(1)$.

(B) $H_n = \Theta(\log \log n)$.

(C) $H_n = \Theta(\sqrt{\log n})$.

(D) $H_n = \Theta(\log n)$.

(E) $H_n = \Theta(\log^2 n)$.

And how much is this?

$$T_n = \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \frac{1}{j}$$

is equal to

(A) $T_n = \Theta(n)$.

(B) $T_n = \Theta(n \log n)$.

(C) $T_n = \Theta(n \log^2 n)$.

(D) $T_n = \Theta(n^2)$.

(E) $T_n = \Theta(n^3)$.

A Slick Analysis of **QuickSort**

Continued...

$$\mathbf{E}\Big[Q(A)\Big] = \sum_{1 \le i < j \le n} \mathbf{E}[X_{ij}] = \sum_{1 \le i < j \le n} \Pr[R_{ij}].$$

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = \sum_{1 \le i < j \le n} \Pr\Big[R_{ij}\Big] = \sum_{1 \le i < j \le n} \frac{2}{j-i+1}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\mathbf{Pr}[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = \sum_{1 \le i < j \le n} \frac{2}{j-i+1}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\mathbf{Pr}[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = \sum_{1\le i<j\le n} \frac{2}{j-i+1}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

A Slick Analysis of QuickSort

Continued...

Lemma

$\mathbf{Pr}[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\left[Q(A)\right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = 2\sum_{i=1}^{n-1}\sum_{i<j}^{n}\frac{1}{j-i+1}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = 2\sum_{i=1}^{n-1}\sum_{i<j}^{n}\frac{1}{j-i+1}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\mathbf{Pr}[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = 2\sum_{i=1}^{n-1}\sum_{i<j}^{n}\frac{1}{j-i+1} \leq 2\sum_{i=1}^{n-1}\quad\sum_{\Delta=2}^{n-i+1}\frac{1}{\Delta}$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = 2\sum_{i=1}^{n-1}\sum_{i<j}^{n}\frac{1}{j-i+1} \leq 2\sum_{i=1}^{n-1}\sum_{\Delta=2}^{n-i+1}\frac{1}{\Delta}$$

$$\leq 2\sum_{i=1}^{n-1}(H_{n-i+1}-1) \leq 2\sum_{1\leq i<n} H_n$$

A Slick Analysis of **QuickSort**

Continued...

Lemma

$\Pr[R_{ij}] = \frac{2}{j-i+1}$.

$$\mathbf{E}\Big[Q(A)\Big] = 2\sum_{i=1}^{n-1}\sum_{i<j}^{n}\frac{1}{j-i+1} \leq 2\sum_{i=1}^{n-1}\sum_{\Delta=2}^{n-i+1}\frac{1}{\Delta}$$

$$\leq 2\sum_{i=1}^{n-1}(H_{n-i+1}-1) \leq 2\sum_{1\leq i<n} H_n$$

$$\leq 2nH_n = O(n\log n)$$

Where do I get random bits?

Question: Are true random bits available in practice?

1. Buy them!
2. CPUs use physical phenomena to generate random bits.
3. Can use pseudo-random bits or semi-random bits from nature. Several fundamental unresolved questions in complexity theory on this topic. Beyond the scope of this course.
4. In practice pseudo-random generators work quite well in many applications.
5. The model is interesting to think in the abstract and is very useful even as a theoretical construct. One can *derandomize* randomized algorithms to obtain deterministic algorithms.