Randomized Algorithms

Andreas Klappenecker

Randomized Algorithms

A randomized algorithm is an algorithm that makes random choices during their execution.

A randomized algorithm uses values generated by a random number generator to decide the next step at several branches of its execution.

Therefore, the steps taken by a randomized algorithm might differ from execution to execution, even if the input remains the same.

Why Randomization?

Randomization can lead to simple algorithms that are easy to implement.

Randomization can lead to efficient implementations.

Running Time

The designer of a randomized algorithm must determine what kind of running time one can expect.

The running time is now a random variable, and one needs tools from probability theory to estimate it.

Motivation (1)

Suppose that a company has several servers containing its database. The database is stored in several locations (e.g. east coast and west coast).

At the end of the business day, the company wants to verify that the copies of the databases are still consistent. Transmission of the data is not feasible. How can we whether the content is the same?

Motivation (2)

Suppose we have implemented an extremely fast algorithm to multiply very large matrices (e.g. of dimension 100,000x100,000).

How can we verify whether the computation was correct?

Motivation (3)

In the RSA key exchange, we need to form the product of two very large primes (each having 1000 digits or more).

How can we efficiently check whether a number is prime?

Basics from Probability Theory

Sample Spaces

The possible outcomes of an experiment are called the sample space Ω.

For example, the sample space of a coin tossing experiment is Ω={head, tail}.

The sample space of rolling a dice is Ω={1,2,3,4,5,6}.

σ-Algebra

- A probability measure is not necessarily defined on all subsets of the sample space, but only on those that are considered events. We will have a uniform way of reasoning about event by requiring that they form a σ-algebra.
- A σ-algebra F is a collection of subsets of a sample space Ω such that
  - the empty set is contained in F,
  - if E in F, then its complement $E^c = \Omega \backslash E$ is in F,
  - a countable union of sets in F is contained in F.

σ-Algebra Example

Let $\Omega = \{1,2,3,4,5,6\}$ the sample space of a die.

Suppose we are interested in the events:

- $D = \{1,2\}$, the value is less than 3.
- $E = \{3,4,5,6\}$, the value is 3 or more.

Then the smallest σ-algebra F containing D and E is given by $F=\{ \varnothing, D, E, \Omega \}$.

The empty set $\varnothing$ is called the impossible event.

The set Ω is called the certain event.

σ-Algebra

The σ-algebra allows one to talk about

- the impossible event
- complementary event
- the union of events
- the certain event

When rolling a dice, the event that the outcome is an even face value is {2,4,6}. The event that the outcome is a value larger than 4 is {5,6}.

Operations on Events

Let D and E be events. Then

- $D \cup E$ is an event
- $D \cap E$ is an event
- $D \setminus E$ is an event

Indeed, let $E_1=D$, $E_2=E$, and $E_3=E_4=\ldots=\varnothing$. Then

$$\bigcup E_i = D \cup E.$$

The other two properties are also easy to show.

Probability Measure

Let F be a σ-algebra over a sample space Ω. A probability measure on F is a function Pr: F -> [0,1] such that

- the certain event satisfies Pr[Ω]=1,
- if the events E_1, E_2, ... in F are mutually disjoint, then

$$Pr[\bigcup_{k=1}^{\infty} E_k] = \sum_{k=1}^{\infty} Pr[E_k]$$

Properties of Probability Measures

Let E be an event. Then

$$1 = \Pr[\Omega] = \Pr[E] + \Pr[E^c],$$

as E and E^c are disjoint.

Therefore, the complementary event E^c has probability

$$\Pr[E^c] = 1 - \Pr[E].$$

In particular, the impossible event has probability

$$\Pr[\varnothing]=1-\Pr[\Omega]=0$$

Properties of Probability Measures

Let D and E be events such that $D \subseteq E$.

Then $Pr[D] <= Pr[E]$.

Why?

Properties of Probability Measures

Let D and E be events. Then

$$\Pr[\, D \cup E \,] = \Pr[D] + \Pr[E] - \Pr[D \cap E].$$

Indeed, we have

(a) $\Pr[D] = \Pr[D - (D \cap E)] + \Pr[D \cap E]$,

(b) $\Pr[E] = \Pr[E - (D \cap E)] + \Pr[D \cap E]$.

Since

$\Pr[D \cup E\,] = \Pr[D - (D \cap E)] + \Pr[E - (D \cap E)] + \Pr[D \cap E]$,

the claim follows from (a) and (b).

Uniform Probability Distribution

Let Ω be a finite sample space.

Let $F = P(\Omega)$ be the σ-algebra consisting of all subsets of Ω.

Then the probability measure $Pr: F \to [0,1]$ defined by

$$Pr[\{s\}] = 1/|\Omega|$$

for all s in Ω is called the uniform probability distribution on Ω.

Continuous Probability Distribution

The continuous uniform probability distribution over an interval [a,b] associates to each subinterval [c,d] of [a,b] the probability

$$\Pr[\,[c,d]\,] = (d-c)/(b-a).$$

Notice that the probability of any event {x} with x in [a,b] is 0, since $\Pr[\,\{x\}\,] = \Pr[\,[x,x]\,] = 0$.

Continuous Probability Distribution

For the sample space $\Omega = [a,b]$, one cannot choose the σ-algebra $F=P(\Omega)$, since there does not exist any function on $P(\Omega) = P([a,b])$ that satisfies our axioms of a probability measure.

Instead, define F to be the smallest σ-algebra on $\Omega = [a,b]$ that contains the intervals [c,d] for all c,d in the range $a <= c <= d <= b$. Then there exists a function Pr: F -> [0,1] such that Pr[[c,d]] = (d-c)/(b-a). It is called the Borel measure on F.

Union Bound

Let I ⊆{1,2,3,...}. Let E_i with i in I be a set of events.

These events do not need to be disjoint.

Then the union bound states that

$$\Pr[\bigcup_{i \in I} E_i] \leq \sum_{i \in I} \Pr[E_i]$$

This simple bound is enormously useful, as it is easy to compute.

Conditional Probabilities

Let D and E be events such that Pr[E] >0.

The conditional probability Pr[D|E] is defined as

$$\Pr[D|E] = \frac{\Pr[D \cap E]}{\Pr[E]}$$

One can interpret Pr[D|E] as the probability that the event D occurs, assuming that the event E occurs.

Useful Multiplication Formula

Quite often, it is easy to determine conditional probabilities:

$$\Pr[D \cap E] = \Pr[D|E] \Pr[E]$$

Independent Events

Two events D and E are called independent if and only if

$$\Pr[D \cap E] = \Pr[D] \Pr[E]$$

If D and E are independent, then

$$\Pr[D|E] = \Pr[D]$$

Verifying Polynomial Identities

Polynomial Identities

Suppose that we want to check whether two polynomials in x with integer coefficients are the same.

For example, is

$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6)$

the same as

x^6-7x^3+25 ?

Deterministic Algorithm

Given two polynomials F(x) and G(x). Convert them to the canonical form

$$\sum_{k=0}^{d} c_k x^k$$

If the canonical forms are the same, then the polynomials F(x) and G(x) must be the same.

Slow! It takes $\Theta(d^2)$ coefficient multiplications.

Randomized Algorithm

Input: Two polynomials $F(x)$ and $G(x)$.

Let $d = \max(\deg F(x), \deg G(x))$.

Choose an integer r uniformly at random from the interval $[1..100d]$.

return $F(r) = G(r)$.

Fast! $O(d)$ multiplications of coefficients.

However, the result can be wrong!

Randomized Algorithm

The algorithm will never err if F(x) is the same as G(x).

The algorithm might err if F(x) and G(x) are not equal.

How likely is it that the algorithm errs?

Randomized Algorithm

The algorithm will report that F and G are the same although they are different if and only if r is a root of the polynomial F(x)-G(x).

However, F(x)-G(x) has degree at most d. There are at most d integers r that can be a root of F(x)-G(x).

Since there are 100d integers in [1..100d], the chance that the algorithm errs is <= d/100d = 1/100.

Reducing the Failure Probability, Version 1

We could reduce the probability of failure by choosing a larger range of integers. However, [1..1000d] will only give Pr[failure] <= 1/1000.

This does not offer too much improvement, so we keep the originally proposed randomized algorithm.

Reducing the Failure Probability, Version 2

Let us run the algorithm k times.

Let E_i denote the event that, on the i-th run of the algorithm, we choose a root r_i such that $F(r_i)-G(r_i)=0$.

The events E_i are independent. Failure probability:

$$\Pr[E_1 \cap E_2 \cap \cdots \cap E_k] = \prod_{i=1}^{k} \Pr[E_i]$$

$$\leq \prod_{i=1}^{k} \frac{d}{100d} = \left(\frac{1}{100}\right)^k$$

Random Variables

Random Variable

Let F be a σ-algebra over a sample space Ω. A random variable X is a function Ω -> R such that

$$\{ z \text{ in } \Omega \mid X(z) <= x \}$$

is an event in F for each x in R.

We write X <= x for this event.

There is nothing random about a random variable!

It is simply a function that allows one to specify events as preimages.

Example 1

Let X be the random variable denoting the sum of face values of a pair of dice. Then X <= 3 is a shorthand for the event {(1,1), (1,2), (2,1)}.

Example 2

Let Y be the random variable counting the number of heads during three subsequent coin tosses. Then

- Y <= 0 is the event {(tail,tail,tail)}
- Y <= 1 is the event {(tail,tail,tail), (head,tail,tail), (tail,head,tail), (tail,tail,head) }

Discrete Random Variable

A random variable with countable image is called a discrete random variable.

For discrete random variables, X=a is an event.

Expectation Value

Let X be a discrete random variable over a probability space (Ω,F,Pr).

The expectation value (or mean) of X is given by

$$\mathrm{E}[X] = \sum_{\alpha \in X(\Omega)} \alpha \Pr[X = \alpha]$$

Example

Let X be the random variable denoting the sum of face values of a pair of fair dice.

What is the expectation value E[X] ?

Example 1

- Let Y denote the number of heads in three subsequent fair coin tosses.
- Y=0 is { (t,t,t) }
- Y=1 is { (h,t,t), (t,h,t), (t,t,h) }
- Y=2 is { (h,h,t), (h,t,h), (t,h,h) }
- Y=3 is { (h,h,h) }
- Pr[Y=0] = Pr[Y=3] = 1/8, Pr[Y=1] =Pr[Y=2]= 3/8

- E[X] = 0(1/8)+1(3/8)+2(3/8)+3(1/8) = 12/8=1.5

Example 2

Two dice totals						
Die 1	Die 2					
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$E[X] = 2Pr[X=2]+3Pr[X=3]+\ldots+12Pr[X=12]$

$= 2(1/36)+3(2/36)+4(3/36)+5(4/36)+6(5/36)+7(6/36)$

$+8(5/36)+9(4/36)+10(3/36)+11(2/36)+12(1/36)$

$= 7$

Linearity of Expectation

Let X and Y be random variables.

Let a and b be real numbers.

Then

$$E[aX+bY] = aE[X]+bE[Y].$$

Simple but extremely useful!

Hat Check Girl

Hat Check Girl

Suppose n persons give their hat to the hat check girl. The girl is upset and hands each person a random hat (where the hat is chosen uniformly at random).

How many persons can expect to get their own hat back?

Hat Check Girl

- The sample space $\Omega = \{1,2,\ldots,n\}$.
- We will allow all subsets of Ω to be events, that is, the σ-algebra is $F=P(\Omega)$.
- For p in Ω, the event {p} has the interpretation that person p received her own hat.
- Since each of the n persons has an equal chance to receive the hat of person p, it follows that Pr[person p receives her own hat] = Pr[{p}]=1/n

Hat Check Girl

- Let X_i denote the random variable that is
  - equal to 1 if the i-th person receives her own hat,
  - and 0 otherwise
- Then $Pr[X_i = 1]=1/n$
- $E[X_i] = 1\ Pr[X_i=1] + 0\ Pr[X_i=0] = 1/n$
- The random variable $X = X_1+ X_2 + ... + X_n$ counts the number of person receiving their own hat.

Hat Check Girl

The number of persons receiving their own hat is expected to be equal to

$$E[X] = n(1/n) = 1$$

by linearity of expectation.

Geometric Random Variables and Coupon Collection

Bernoulli Distribution or Biased Coins

Bernoulli Distribution. Tossing a biased coin can be described by a random variable X that takes the value 1 if the outcome of the experiment is **head**, and the value 0 if the outcome is **tail**. Assume that $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$ for some real number $p \in (0, 1)$. The random variable X is said to have the Bernoulli distribution with parameter p. We can compute the expectation value and the variance as follows:

$$\mathrm{E}[X] = p, \qquad \mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = p - p^2 = p(1 - p).$$

Geometric Distribution

Geometric Distribution. Suppose we keep tossing a biased coin, which has the Bernoulli distribution with parameter p, until the event **head** occurs. Let the random variable X denote the number of coin flips needed in this experiment. We say that X is geometrically distributed with parameter p. The density function of X is given by

$$p_X(x) = \Pr[X = x] = p(1-p)^{x-1}$$

for $x = 1, 2, \ldots,$ and $p_X(x) = 0$ otherwise. The expectation value and the variance of X are given by

$$\mathrm{E}[X] = \frac{1}{p}, \qquad \mathrm{Var}[X] = \frac{1-p}{p^2}.$$

Coupon Collection

The Coupon Collector Problem. The hat check girl is a compulsive coupon collector. Currently, she is collecting charming Harry Potter characters that are contained in overpriced serial boxes. There are n different characters, and each box contains one character. She wants to get the complete set. How many boxes does she have to buy, on average, to obtain one complete collection?

Coupon Collection

Let X denote the random variable counting the number of boxes required to collect at least one character of each type. Our goal is to determine $\mathrm{E}[X]$. Let X_k denote the random variables counting the number of boxes that the hat check girl buy to get the $(k+1)$-th character, after she has already collected k characters. The probability to draw one of the remaining characters is $p_k = (n-k)/n$. Hence X_k is a geometrically distributed random variable with parameter p_k. Consequently, $\mathrm{E}[X_k] = 1/p_k = n/(n-k)$.

The random variable X is the given by the sum $X = \sum_{k=0}^{n-1} X_k$. Linearity of expectation shows that

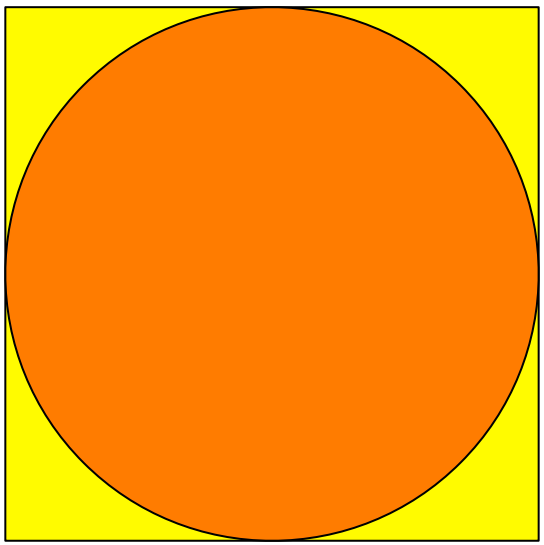
$$\mathrm{E}[X] = \sum_{k=0}^{n-1} \mathrm{E}[X_k] = \sum_{k=0}^{n-1} \frac{n}{n-k} = n \sum_{k=1}^{n} \frac{1}{k} = nH_n.$$

The Monte Carlo Method

Motivation

How can we calculate π ?

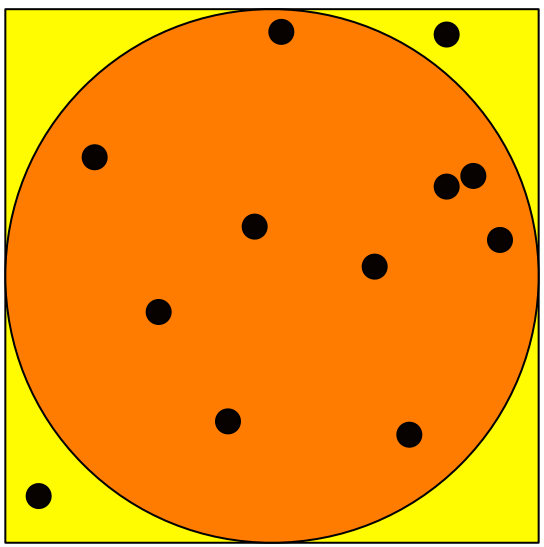
The Main Idea (1)



Given: A 2x2 square centered at (0,0) with a circle of radius 1 inscribed.

The area of the circle is π and the area of the square is 4.

The Main Idea (2)



Choose points uniformly at random in the square.

Ratio of (# points in the circle)/(# all points) approximates $\pi/4$.

Randomly Chosen Points

- Let (X,Y) be a point chosen uniformly at random in the 2x2 square. Equivalently, we can choose X and Y independently from a uniform distribution on [-1,1].

- Let Z be the indicator random variable that is 1 if the point falls within the circle and 0 otherwise. Put differently, Z=1 if and only if $\sqrt{X^2 + Y^2} \leq 1$

- Pr[Z=1] = π/4.

Estimating π

- Suppose we repeat this experiment m times, where Z_i denotes the value of Z at the i-th run.
- Let $W = \sum_{i=1}^{m} Z_i$
- Then

$$E[W] = E[\sum_{i=1}^{m} Z_i] = \sum_{i=1}^{m} E[Z_i] = m\pi/4.$$

- Therefore, W'=4W/m is a natural estimate for π.

A Chernoff Bound

- Let $X_1, X_2, \ldots, X_m$ be random variables such that $\Pr[X_i=1]=p_i$ and $\Pr[X_i=0]=1-p_i$.
- Let $X = \sum_{i=1}^{m} X_i$
- For $0<\delta<1$, we have

$$\Pr[|X - E[X]| \geq \delta E[X]] \leq 2\exp(-E[X]\delta^2/3)$$

Approximating π

Applying the Chernoff bound to our estimate of π, we get

$$\begin{aligned}
\Pr[|W' - \pi| \geq \epsilon\pi] &= \Pr\left[\left|W - \tfrac{m\pi}{4}\right| \geq \tfrac{\epsilon m\pi}{4}\right] \\
&= \Pr[|W - E[W]| \geq \epsilon E[W]] \\
&\leq 2\exp(-\epsilon^2 E[W]/3) \\
&= 2\exp(-m\pi\epsilon^2/12)
\end{aligned}$$

Therefore, we see that the probability W′ deviates significantly from π exponentially decreases with the number of trials m.