Probabilistic Analysis of Deterministic Algorithms

Also called *average case analysis*.

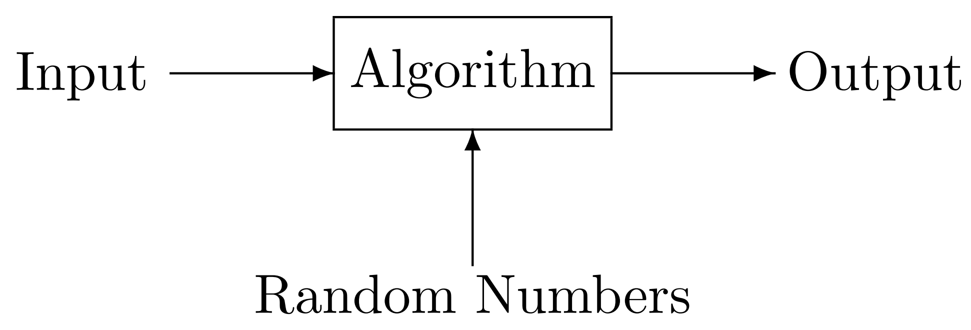
I **:** Input instance.

$T_A(I)$ **:** Time taken on instance I by algorithm A.

$\Pr(I)$ **:** Probability that instance I arises in practice.

Expected time taken by algorithm A: $\sum_I T_A(I) \Pr(I)$

Randomized Algorithms



Effect of Random Numbers:

- Different answers may be generated. *(Each not necessarily correct.)*
- Time taken to generate may be different.

Hope:

- Usually a correct answer will be generated.
- Usually, time taken will be small. Usually time taken will be less than good deterministic algorithms.

Evaluating a Randomized Algorithm

I **:** Input instance.

R **:** Random numbers given.

$T_{I,R}$ **:** Time taken for input I and Random numbers R

Expected time for input I: $\sum_R T(I,R)\Pr(R)$

Measure 1: (Worst) Expected time

$$\max_I \left(\sum_R T(I,R)\Pr(R) \right)$$

Measure 2: "High Probability Analysis"

More detailed than Expectation.

"What is the probability that time > ...?"

Formal definition later.

Probability Theory Refresher

Probability Space

Set of events which are (elementary) outcomes of an "experiment". Also *Sample Space*.

Experiment	Probability Space $\mathcal{S}$
Flipping 2 coins	{tt,th,ht,hh}
Picking a card	$\{\clubsuit 2, \clubsuit 3, \ldots, \spadesuit A\}$
Permutation Routing on 4 node hypercube	(Intermediate destinations) $\{(0,0,0,0),(0,0,0,1),\ldots,$ $\ldots,(2,3,3,3),(3,3,3,3)\}$

Probability Distribution: Function Pr from $\mathcal{S}$ to Real numbers s.t.

1. For any $s \in \mathcal{S}$, $\Pr(s) \geq 0$.
2. $\sum_{s \in \mathcal{S}} Pr(s) = 1$

Examples:

$\Pr(tt) = \Pr(th) = \Pr(ht) = Pr(hh) = \frac{1}{4}$.

$\Pr(\text{each card}) = \frac{1}{52}$

$\Pr(\text{Each choice of intermediate destinations}) = \frac{1}{64}$

Non Elementary Events

Subsets of the probability space.

Examples:

1. First coin out of two tosses is a head $= \{ht, hh\}$.
2. An ace is drawn when a card is chosen from a deck $= \{\spadesuit A, \heartsuit A, \diamondsuit A, \clubsuit A\}$
3. Some node has all 4 packets at end of phase 1 $= \{(0,0,0,0), (1,1,1,1), (2,2,2,2), (3,3,3,3)\}$.

Probability of non elementary events: Sum of the probabilities of the events in the associated subset.

Probabilities for the above events: $\frac{1}{2}, \frac{1}{13}, \frac{1}{16}$

Random Variable

Function from probability space to R.

Examples:

H_2 **:** Number of heads in two coin flips.

$\mathcal{S} = \{tt, th, ht, hh\}$

$H_2(tt) = 0$

$H_2(th) = 1$

$H_2(ht) = 1$

$H_2(hh) = 2$

More customary: $H_2 = 0$ when tt, 1 when th or ht, and 2 when hh.

Q **:** Time required by Quicksort when input is a random permutation.

T **:** Time to deliver packets in a certain network when each processor sends a packet to a randomly chosen destination.

Expectation of a Random Variable

X : random variable over probability space $\mathcal{S}$

$$E[X] \equiv \sum_x x \Pr(X = x)$$

or alternatively

$$E[X] = \sum_{s \in \mathcal{S}} \Pr(s) X(s)$$

Example:
$E[H_2]$
$= 0 \cdot \Pr[H_2 = 0] + 1 \cdot \Pr[H_2 = 1] + 2 \cdot \Pr[H_2 = 2]$
$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4}$
$= 1$

Alternatively:
$E[H_2]$
$= \frac{1}{4} X(tt) + \frac{1}{4} X(th) + \frac{1}{4} X(ht) + \frac{1}{4} X(hh)$
$= \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 2$
$= 1$

Bernoulli Random Variables

Random variables taking values 0 or 1.

Example: A_{lp} = Number of times a packet p will cross a link l assuming its path is simple.

Non Example: H_2, C_l, T_π.

Expectation of a Bernoulli random variable:

X **:** $\Pr(X = 1) = p$, $\Pr(X = 0) = 1 - p$.

$E[X] \ = \sum_x \Pr(X = x)$

$\ = 0 \cdot \Pr(X = 0) + 1 \cdot \Pr(X = 1)$

i.e. $E[X] = p$

Typical Computation in Probability Theory

Probability of an event:

X = Gambler doubles his money, ...
Y = Time to deliver all packets > 100
Z = Time to deliver given packet > 100

Upper/lower bounds may be acceptable.

Expectation of a random variable:

A = Winnings of the gambler
B = Time to deliver all packets
C = Time to deliver given packet

Upper/lower bounds may be acceptable.

How to compute probability/expectation: Understand (a) structure of the events/random variables, (b) relationship between expectation and probabilities.

Examples of Event Structure

UNION EVENT:
Suppose $A = A_1 \cup A_2 \cup \ldots \cup A_k$. Then

$$\Pr[A] \;\leq\; \sum_{i=1}^{i=k} \Pr[A_i]$$

Proof: Venn Diagram.

Example:
A_i = Ace of spades drawn in ith trial. Pr = 1/52
Pr[at least 1 ace in 2 trials]
$\leq \Pr[A_1] + \Pr[A_2] = 1/26$
Reasonably accurate (within $1/52^2$).

Very useful if not much overlap among A_i.

SUBSET EVENT
$A \subseteq B \Rightarrow \Pr[A] \leq \Pr[B]$

COMPLEMENT EVENT
A is the complement of $B \Rightarrow \Pr[A] = 1 - \Pr[B]$

Independence

Events X and Y are independent if

$$\Pr[X \cap Y] = \Pr[X] \cdot Pr[Y]$$

Intuition: Knowing X has happened does not help in predicting whether Y also happens. Example: X = head on first toss of a balanced coin, Y = head on second toss of a balanced coin.

Random variables X and Y are independent if for all real numbers x and y,

$$\Pr(X = x \text{ and } Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

Intuition: Knowing the value of X doesn't help us predict the value of Y.

Structure in Random Variables

X, Y, Z random variables on $\mathcal{S}$.

Definition: $Z = X + Y$ iff $Z(s) = X(s) + Y(s)$ for all $s \in \mathcal{S}$.

Lemma (Linearity of Expectation): $Z = X + Y$ then $E[Z] = E[X] + E[Y]$.

Example: X = Number of heads in first 10 coin tosses. Y = Number of heads in next 10. Z = number in the first 20.

Definition: $Z \geq X$ iff $Z(s) \geq X(s)$ for all $s \in \mathcal{S}$.

Lemma: $Z \geq X$ then $E[Z] \geq E[X]$

Proofs: Exercise.

Expectation vs. Probability

The central idea: A random variable takes values *too far away* from its expectation with *low* probability.

- Markov's inequality: relates the probability of being far from the expectation. Useful even when we do not know much about the structure of the variable.
- Chernoff bounds: relates the probability of being far from the expectation. But the variable must be a sum of independent Bernoulli random variables.

Obviously, Chernoff bounds are sharper than those given by Markov's inequality. However, Markov's inequality is more applicable.

Markov's Inequality

Theorem: If a random variable X only takes non-negative values, then

$$P(X > k) \leq \frac{E[X]}{k}$$

Example:
E[number of heads in 100 tosses] $= 50$
Pr[$\geq$75 heads in 100 tosses] $\leq \frac{50}{75} = \frac{2}{3}$

...Stronger bounds possible

Chernoff Bounds

Theorem: Let $X_1, \ldots, X_n$ be independent Bernoulli Random variables with $\Pr[X_i = 1] = p_i$.
Let $X = X_1 + \cdots + X_n$.
Let $\mu = E[X] = \sum_i E[X_i] = \sum_i p_i$.
Then

$$\Pr[X \geq \beta\mu] \leq e^{(1-\frac{1}{\beta}-\ln\beta)\beta\mu} \leq \left(\frac{\beta}{e}\right)^{-\beta\mu} \qquad \text{for } \beta \geq 0$$

$$\Pr[X \geq m] \leq \left(\frac{m}{\mu e}\right)^{-m} \qquad \text{for } m \leq 0$$

$$\Pr[X \geq (1+\epsilon)\mu] \leq e^{-\epsilon^2\mu/3} \qquad \text{for } 0 < \epsilon < 1$$

$$\Pr[X \geq (1+\epsilon)\mu] \leq e^{-\epsilon^2\mu/4} \qquad \text{for } 0 < \epsilon < 2e - 1$$

$$\Pr[X \geq (1+\epsilon)\mu] \leq 2^{-(1+\epsilon)\mu} \qquad \text{for } 2e - 1 \leq \epsilon$$

$$\Pr[X \leq (1-\epsilon)\mu] \leq e^{-\epsilon^2\mu/2} \qquad \text{for } 0 < \epsilon$$

Example:

E[number of heads in 100 tosses] $= 50$

Pr[$\geq$75 heads in 100 tosses] $\leq e^{-(0.5)^2 50/3} = 0.015$

Identical X_i

$p = p_1 = p_2 = \cdots = p_n$
$\mu = np$
Proof:
Probability that at least m variables out of n take value 1
$\leq$ number of ways of selecting m variables from n
* Probability that given set takes value 1.
$= \binom{n}{m} p^m \leq \left(\frac{npe}{m}\right)^m = \left(\frac{\mu e}{m}\right)^m$

Useful Inequality: $\binom{n}{m} \leq \left(\frac{ne}{m}\right)^m$

“High Probability”

N : Problem size.
$f(N)$: Random variable being studied.
g : function of one variable.
We say “$f(N) = O(g(N))$ with high probability” if for every k there exist constants c, N_0 such that

$$\Pr[f(N) \geq cg(N)] \leq N^{-k}$$

whenever $N \geq N_0$.

Alternative Definition

N : Problem size.
$f(N)$: Random variable being studied.
g : function of one variable.

We say "$f(N) = O(g(N))$ with high probability" if there exist a function h and constant N_0 such that for any k:

$$\Pr[f(N) \geq h(k)g(N)] \leq N^{-k}$$

whenever $N \geq N_0$.

Compositional Properties:

Let $A_i(N) = O(g(N))$ w.h.p. for $i = 1 \ldots m$.
m is at most polynomially large in N.
Let

$A_{\text{series}}(N) = \sum_i A_i(N), \quad A_{\text{parallel}}(N) = \max_i A_i(N)$

Then

$A_{\text{series}}(N) = O(mg(N)), \quad A_{\text{parallel}}(N) = O(g(n))$

Proof: for $A = A_{\text{series}}$:

We know $\Pr[A_i \geq h_i(k)g(N)] \leq N^{-k}$ for $N \geq N_{i0}$

$\Pr[A \geq cmg(N)] \leq \sum_i \Pr[A_i \geq cg(N)]$

If $c = H(k) = \max_i h_i(k), \quad N_0 = \max_i N_{i0}$

Then $\Pr[A_i \geq cg(N)] \leq N^{-k}$

Thus $\Pr[A \geq H(k)mg(N)] \leq mN^{-k} = N^{-k+\log m}$

Thus $\Pr[A \geq H(k' - \log m)mg(N)] \leq N^{-k'}$

Choose $h(k') = H(k' - \log m)$
and note $m = poly(N) \Rightarrow \log m = O(1)$.

Exercises

1. I give a 100 mark multiple choice examination. Each question has 4 alternatives. Each correct answers fetches me 3 marks. Each wrong answer gets me -1 marks. What is the expected number of marks I get? Give a bound on the probability that I get 25 marks more than the expected number. Use Chernoff bounds.

2. I randomly choose $\sqrt{N}$ numbers out of N (with replacement), and determine their median. We would like to use this median as an estimate of the median of the N numbers. Discuss in what way this is likely to be a good estimate.