Probability Review I

This part is from https://www.cs.cmu.edu/~epxing/Class/10701-12f/recitation/Probability_Review.ppt

Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- The big picture
- Examples

Sample space and Events

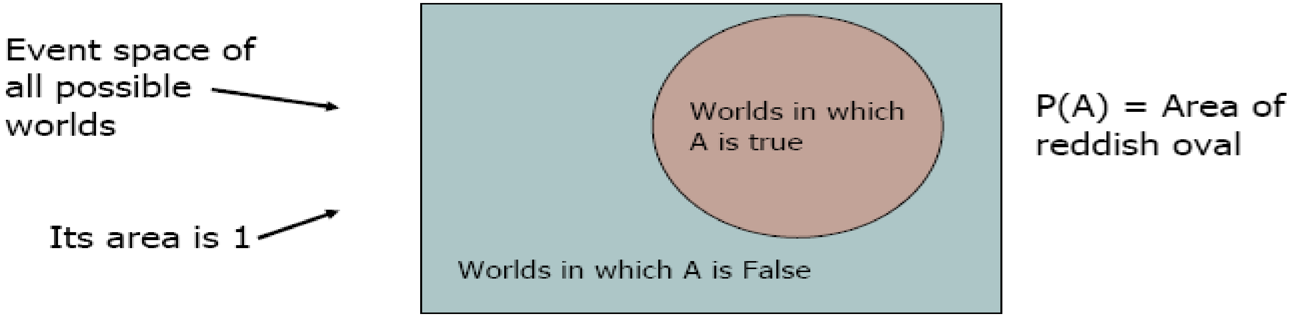
- Ω : Sample Space, result of an experiment
  - If you toss a coin twice $\Omega = \{HH, HT, TH, TT\}$
- Event: a subset of Ω
  - First toss is head = {HH,HT}
- S: event space, a set of events:
  - Closed under finite union and complements
    - Entails other binary operation: union, diff, etc.
  - Contains the empty event and Ω

Probability Measure

- Defined over (Ω, S) s.t.
  - $P(\alpha) >= 0$ for all α in S
  - $P(\Omega) = 1$
  - If α, β are disjoint, then
    - $P(\alpha \text{ U } \beta) = p(\alpha) + p(\beta)$
- We can deduce other axioms from the above ones
  - Ex: $P(\alpha \text{ U } \beta)$ for non-disjoint event

  $P(\alpha \text{ U } \beta) = p(\alpha) + p(\beta) - p(\alpha \cap \beta)$

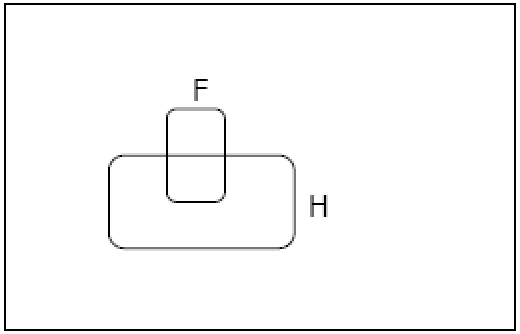
Visualization



- We can go on and define conditional probability, using the above visualization

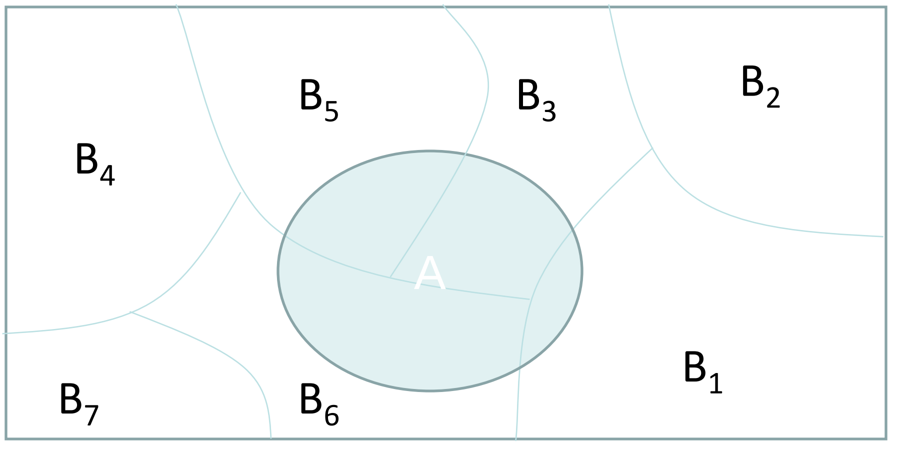
Conditional Probability

P(F|H) = Fraction of worlds in which H is true that also have F true



$$p(f \mid h) = \frac{p(F \cap H)}{p(H)}$$

Rule of total probability

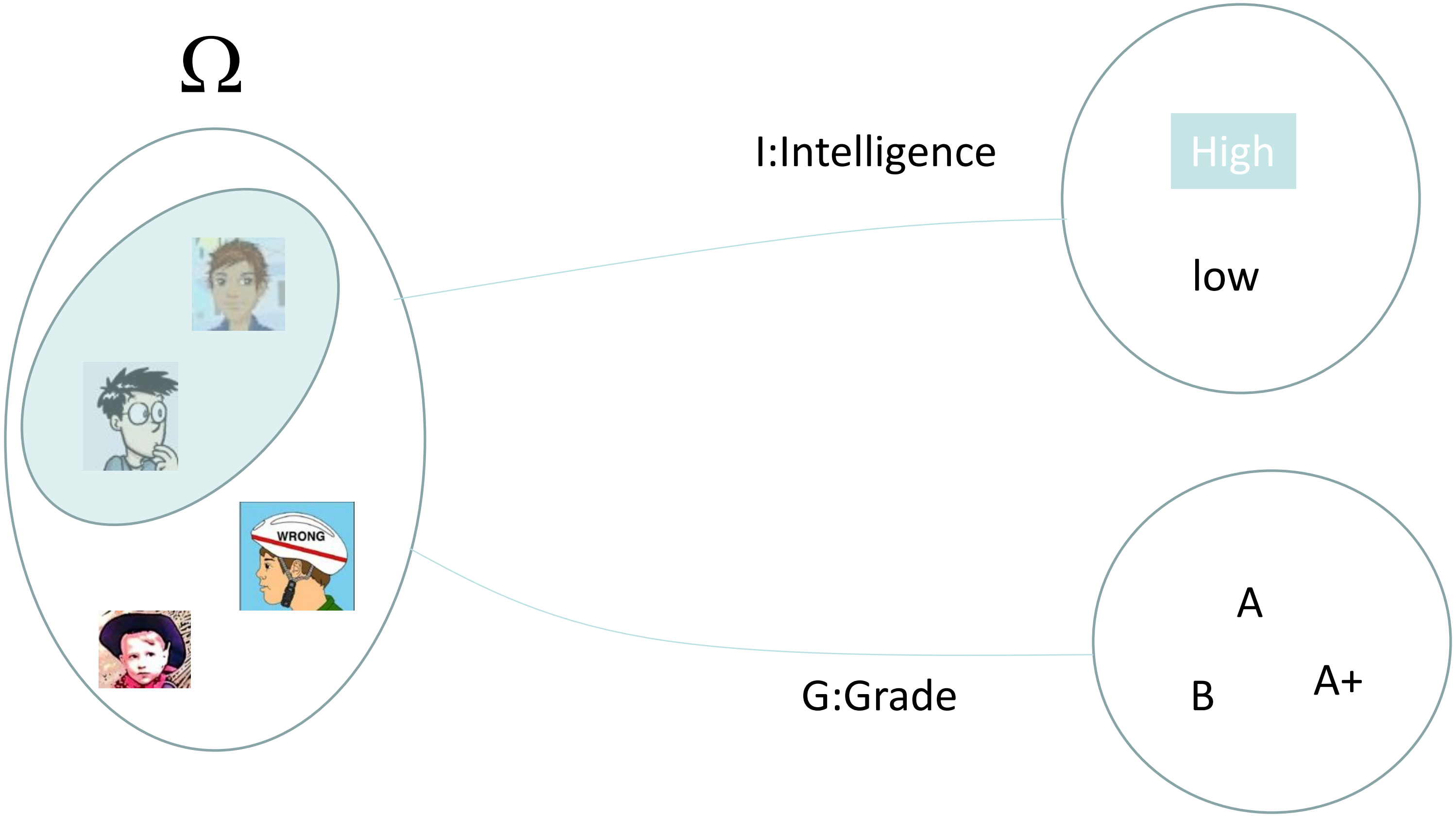


$$p(A) = \sum P(B_i) P(A \mid B_i)$$

From Events to Random Variable

- Almost all the semester we will be dealing with RV
- Concise way of specifying attributes of outcomes
- Modeling students (Grade and Intelligence):
  - $\Omega =$ all possible students
  - What are events
    - Grade_A = all students with grade A
    - Grade_B = all students with grade B
    - Intelligence_High = … with high intelligence
  - Very cumbersome
  - We need “functions” that maps from Ω to an attribute space.
  - P(G = A) = P({student ϵ Ω : G(student) = A})

Random Variables



P(I = high) = P({all students whose intelligence is high})

Discrete Random Variables

- Random variables (RVs) which may take on only a **countable** number of **distinct** values
  - E.g. the total number of tails X you get if you flip 100 coins
- X is a RV with arity *k* if it can take on exactly one value out of $\{x_1, \ldots, x_k\}$
  - E.g. the possible values that X can take on are 0, 1, 2, …, 100

Probability of Discrete RV

- Probability mass function (pmf): $P(X = x_i)$
- Easy facts about pmf
  - $\Sigma_i P(X = x_i) = 1$
  - $P(X = x_i \cap X = x_j) = 0$ if $i \neq j$
  - $P(X = x_i \cup X = x_j) = P(X = x_i) + P(X = x_j)$ if $i \neq j$
  - $P(X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k) = 1$

Common Distributions

- Uniform X   *U*[1, …, *N*]
  - X takes values 1, 2, … *N*
  - P(X = *i*) = 1/*N*
  - E.g. picking balls of different colors from a box
- Binomial X   *Bin*(*n*, *p*)
  - X takes values 0, 1, …, *n*
  - $p(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$
  - E.g. coin flips

Continuous Random Variables

- Probability density function (pdf) instead of probability mass function (pmf)
- A pdf is any function $f(x)$ that describes the probability density in terms of the input variable x.

Probability of Continuous RV

- Properties of pdf
  - $f(x) \geq 0, \forall x$
  - $\int_{-\infty}^{+\infty} f(x) = 1$
- Actual probability can be obtained by taking the integral of pdf
  - E.g. the probability of X being between 0 and 1 is

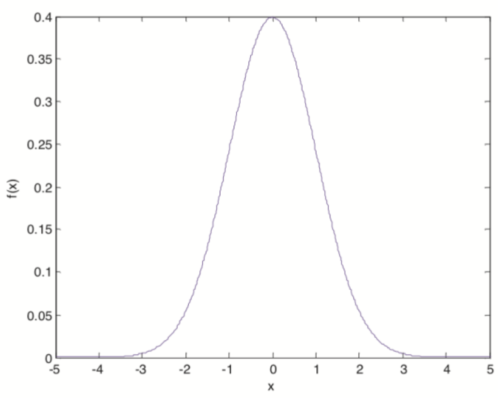
$$P(0 \leq X \leq 1) = \int_0^1 f(x)dx$$

Cumulative Distribution Function

- $F_X(v) = P(X \leq v)$
- Discrete RVs
  - $F_X(v) = \Sigma_{vi} P(X = v_i)$
- Continuous RVs
  - $$F_X(v) = \int_{-\infty}^{v} f(x)dx$$
  - $$\frac{d}{dx} F_x(x) = f(x)$$

Common Distributions

- Normal X $N(\mu, \sigma^2)$
  - $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
  - E.g. the height of the entire population



Probability Review II

This part is from https://www.cs.cmu.edu/~epxing/Class/10701-12f/recitation/Probability_Review.ppt

Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- The big picture
- Examples

Joint Probability Distribution

- Random variables encodes attributes
- Not all possible combination of attributes are equally likely
  - Joint probability distributions quantify this
- P(X= x, Y= y) = P(x, y)
  - Generalizes to N-RVs
  - $\sum_x \sum_y P(X = x, Y = y) = 1$
  - $\int_x \int_y f_{X,Y}(x, y) dx dy = 1$

Chain Rule

- Always true
  - $P(x, y, z) = p(x)\, p(y|x)\, p(z|x, y)$

    $= p(z)\, p(y|z)\, p(x|y, z)$

    $= \ldots$

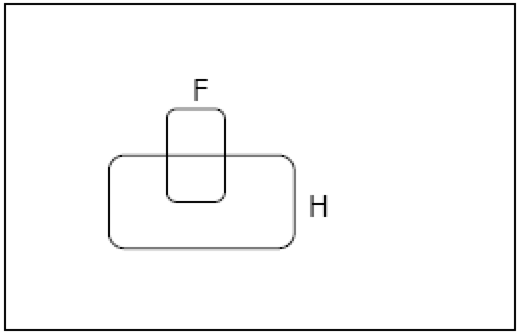
Conditional Probability

events

$$P\left(X = x \middle| Y = y\right) = \frac{P\left(X = x \cap Y = y\right)}{P\left(Y = y\right)}$$

But we will always write it this way:

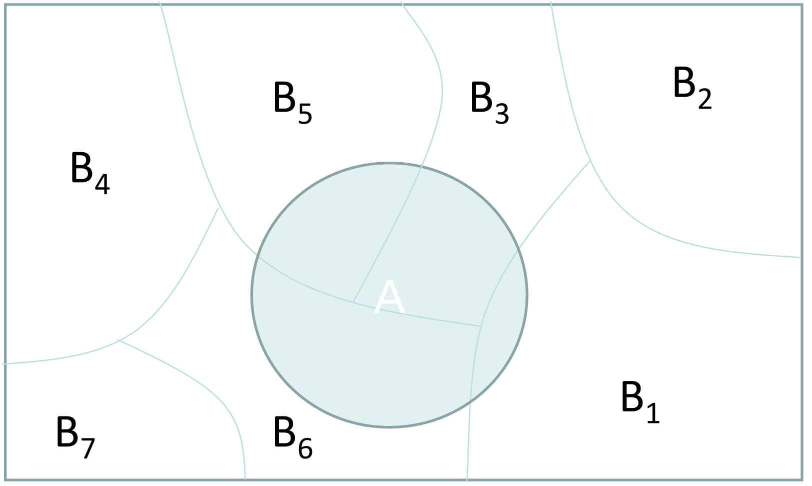
$$P(x \mid y) = \frac{p(x, y)}{p(y)}$$



Marginalization

- We know p(X, Y), what is P(X=x)?
- We can use the low of total probability, why?

$$p(x) = \sum_y P(x, y)$$
$$= \sum_y P(y) P(x \mid y)$$



Marginalization Cont.

- Another example

$$p(x) = \sum_{y,z} P(x, y, z)$$

$$= \sum_{z,y} P(y, z) P(x \mid y, z)$$

Bayes Rule

- We know that P(rain) = 0.5
  - If we also know that the grass is wet, then how this affects our belief about whether it rains or not?

$$P(rain \mid wet) = \frac{P(rain)P(wet \mid rain)}{P(wet)}$$

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(y)}$$

Bayes Rule cont.

- You can condition on more variables

$$P(x \mid y, z) = \frac{P(x \mid z) P(y \mid x, z)}{P(y \mid z)}$$

Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- The big picture
- Examples

Independence

- X is independent of Y means that knowing Y does not change our belief about X.
  - $P(X|Y=y) = P(X)$
  - $P(X=x, Y=y) = P(X=x) P(Y=y)$
  - The above should hold for all x, y
  - It is symmetric and written as $X \perp Y$

Independence

- $X_1, \ldots, X_n$ are independent if and only if

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P\left(X_i \in A_i\right)$$

- If $X_1, \ldots, X_n$ are independent and identically distributed we say they are *iid* (or that they are a random sample) and we write

$$X_1, \ldots, X_n \sim P$$

CI: Conditional Independence

- RV are rarely independent but we can still leverage local structural properties like Conditional Independence.
- $X \perp Y \mid Z$ if once Z is observed, knowing the value of Y does not change our belief about X
  - P(rain $\perp$ sprinkler's on | cloudy)
  - P(rain $\not\perp$ sprinkler's on | wet grass)

Conditional Independence

- $P(X=x \mid Z=z, Y=y) = P(X=x \mid Z=z)$
- $P(Y=y \mid Z=z, X=x) = P(Y=y \mid Z=z)$
- $P(X=x, Y=y \mid Z=z) = P(X=x \mid Z=z)\, P(Y=y \mid Z=z)$

We call these factors : very useful concept !!

Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- Examples

Mean and Variance

- Mean (Expectation): $\mu = E(\mathrm{X})$
  - Discrete RVs: $E(\mathrm{X}) = \sum_{v_i} v_i \mathrm{P}(\mathrm{X} = v_i)$

$$E(g(X)) = \sum_{v_i} g(v_i) P(X = v_i)$$

  - Continuous RVs:

$$E(\mathrm{X}) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Mean and Variance

- Variance: $Var(X) = E((X-\mu)^2)$

  $$Var(X) = E(X^2) - \mu^2$$

  – Discrete RVs: $V(\mathrm{X}) = \sum_{v_i} (v_i - \mu)^2 \, \mathrm{P}(\mathrm{X} = v_i)$

  – Continuous RVs: $V(\mathrm{X}) = \int_{-\infty}^{+\infty} (x-\mu)^2 \, f(x) dx$

- Covariance:

$$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y)) = E(XY) - \mu_x \mu_y$$

Mean and Variance

- Correlation:

$$\rho(X,Y) = Cov(X,Y)/\sigma_x\sigma_y$$

$$-1 \leq \rho(X,Y) \leq 1$$

Properties

- Mean
  - $E(\mathrm{X}+\mathrm{Y}) = E(\mathrm{X}) + E(\mathrm{Y})$
  - $E(a\mathrm{X}) = aE(\mathrm{X})$
  - If X and Y are independent, $E(\mathrm{XY}) = E(\mathrm{X}) \cdot E(\mathrm{Y})$
- Variance
  - $V(a\mathrm{X}+b) = a^2 V(\mathrm{X})$
  - If X and Y are independent, $V(\mathrm{X}+\mathrm{Y}) = V(\mathrm{X}) + V(\mathrm{Y})$

Some more properties

- The conditional expectation of Y given X when the value of X = x is:

$$E\left(Y \mid X = x\right) = \int y * p(y \mid x) dy$$

- The Law of Total Expectation or Law of Iterated Expectation:

$$E(Y) = E\left[E(Y \mid X)\right] = \int E(Y \mid X = x) p_X(x) dx$$

Some more properties

- The law of Total Variance:

$$Var(Y) = Var\left[E(Y \mid X)\right] + E\left[Var(Y \mid X)\right]$$

Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- Examples

Probability Review

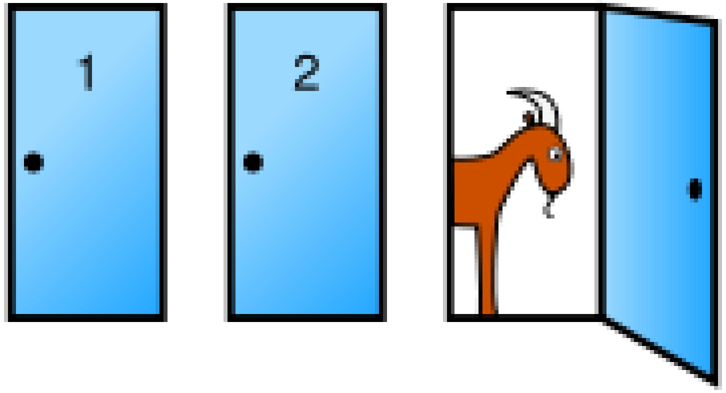
- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- The big picture
- Examples

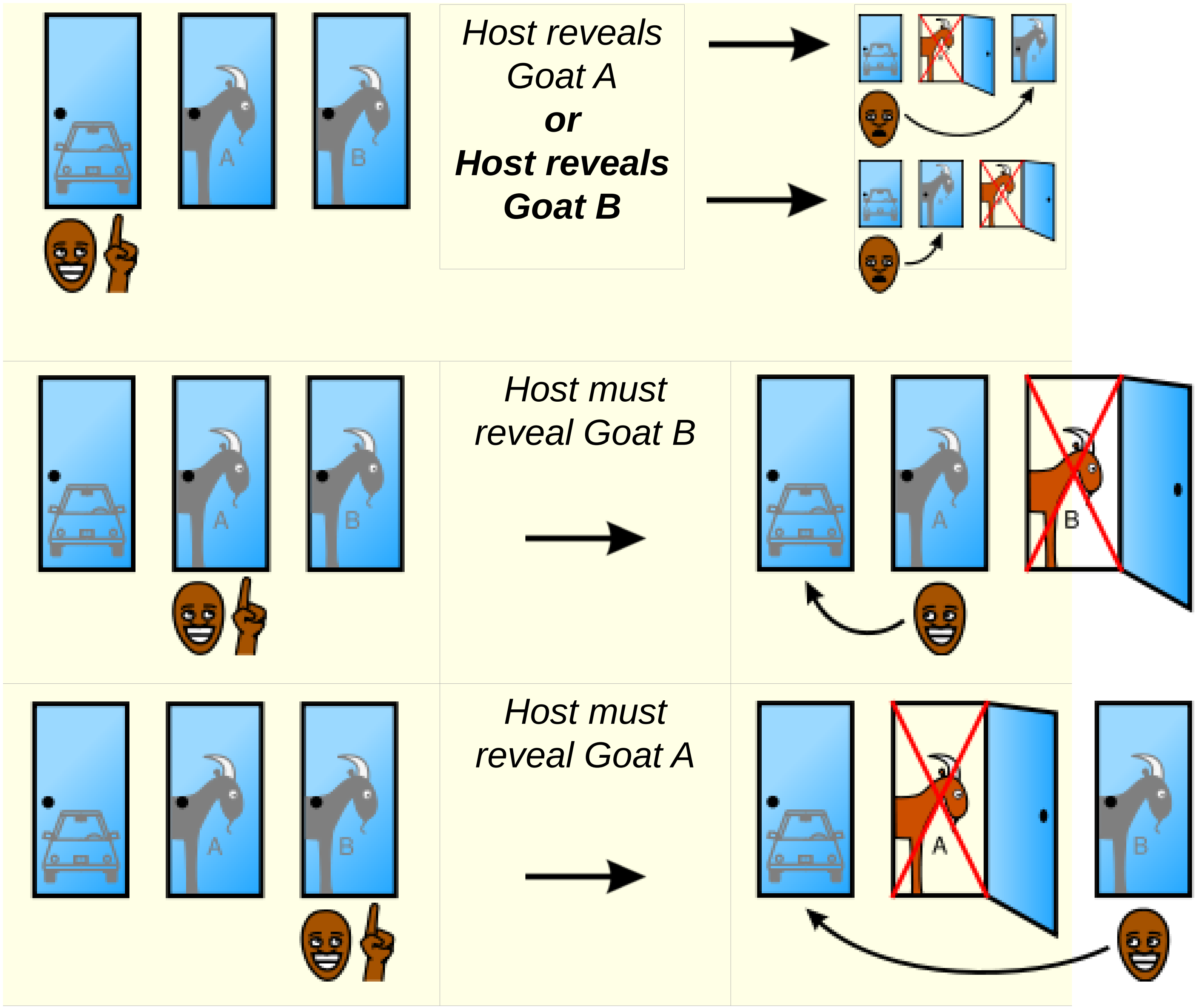
Probability Review III

This part is from https://www.cs.cmu.edu/~epxing/Class/10701-12f/recitation/Probability_Review.ppt

Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?





Host reveals
Goat A
or
Host reveals
Goat B
A
B
Host must
reveal Goat B
Host must
reveal Goat A

Monty Hall Problem: Bayes Rule

- C_i : the car is behind door *i*, *i* = 1, 2, 3
- $P(C_i) = 1/3$
- H_{ij} : the host opens door *j* after you pick door *i*
- $$P(H_{ij} | C_k) = \begin{cases} 0 & i = j \\ 0 & j = k \\ 1/2 & i = k \\ 1 & i \neq k, j \neq k \end{cases}$$

Monty Hall Problem: Bayes Rule cont.

- Without loss of generality, *i*=1, *j*=3
- $$P\left(C_1 \middle| H_{13}\right) = \frac{P\left(H_{13} \middle| C_1\right) P\left(C_1\right)}{P\left(H_{13}\right)}$$
- $$P\left(H_{13} \middle| C_1\right) P\left(C_1\right) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Monty Hall Problem: Bayes Rule cont.

- $$P\left(H_{13}\right)=P\left(H_{13},C_1\right)+P\left(H_{13},C_2\right)+P\left(H_{13},C_3\right)$$
$$=P\left(H_{13}\middle|C_1\right)P\left(C_1\right)+P\left(H_{13}\middle|C_2\right)P\left(C_2\right)$$
$$=\frac{1}{6}+1\cdot\frac{1}{3}$$
$$=\frac{1}{2}$$
- $$P\left(C_1\middle|H_{13}\right)=\frac{1/6}{1/2}=\frac{1}{3}$$

Monty Hall Problem: Bayes Rule cont.

- $$P\left(C_1 \middle| H_{13}\right) = \frac{1/6}{1/2} = \frac{1}{3}$$
- $$P\left(C_2 \middle| H_{13}\right) = 1 - \frac{1}{3} = \frac{2}{3} > P\left(C_1 \middle| H_{13}\right)$$
- *You should switch!*

Acknowledgment

- Carlos Guestrin recitation slides: http://www.cs.cmu.edu/~guestrin/Class/10708/recitations/r1/Probability_and_Statistics_Review.ppt
- Andrew Moore Tutorial: http://www.autonlab.org/tutorials/prob.html
- Monty hall problem: http://en.wikipedia.org/wiki/Monty_Hall_problem
- http://www.cs.cmu.edu/~guestrin/Class/10701-F07/recitation_schedule.html
- Chi-square test for independence

  http://stattrek.com/chi-square-test/independence.aspx

Introduction to Randomized Algorithms

This part of slides come from Srikrishnan Divakaran DA-IICT

Outline

- Preliminaries and Motivation
- Analysis of
  - Randomized Quick Sort
  - Karger’s Min-cut Algorithm
- Basic Analytical Tools

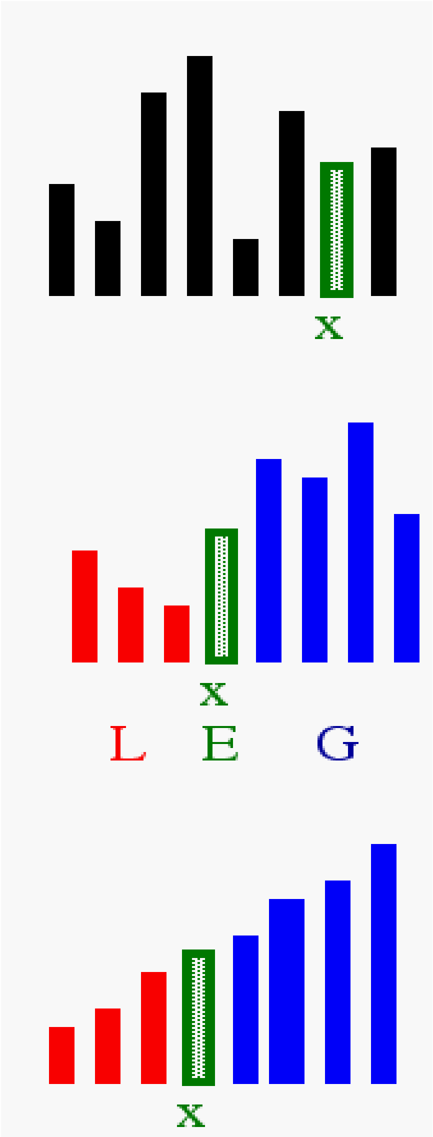
Preliminaries and Motivation

Quick Sort

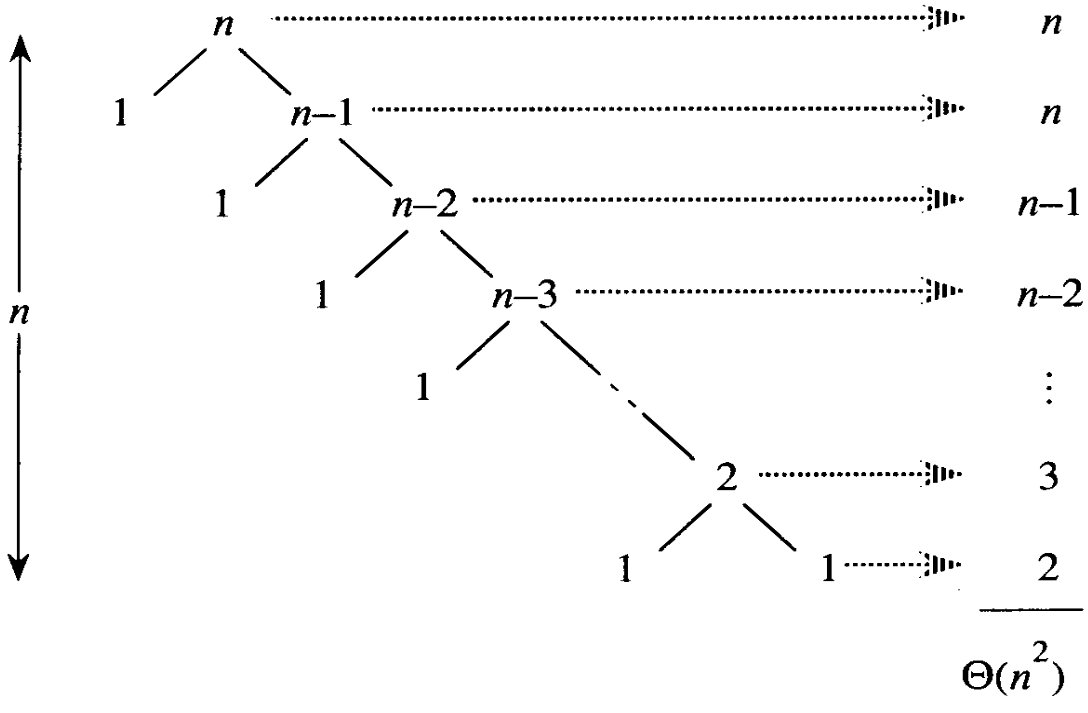
Select: pick an arbitrary element x in S to be the pivot.

Partition: rearrange elements so that elements with value less than x go to List L to the left of x and elements with value greater than x go to the List R to the right of x.

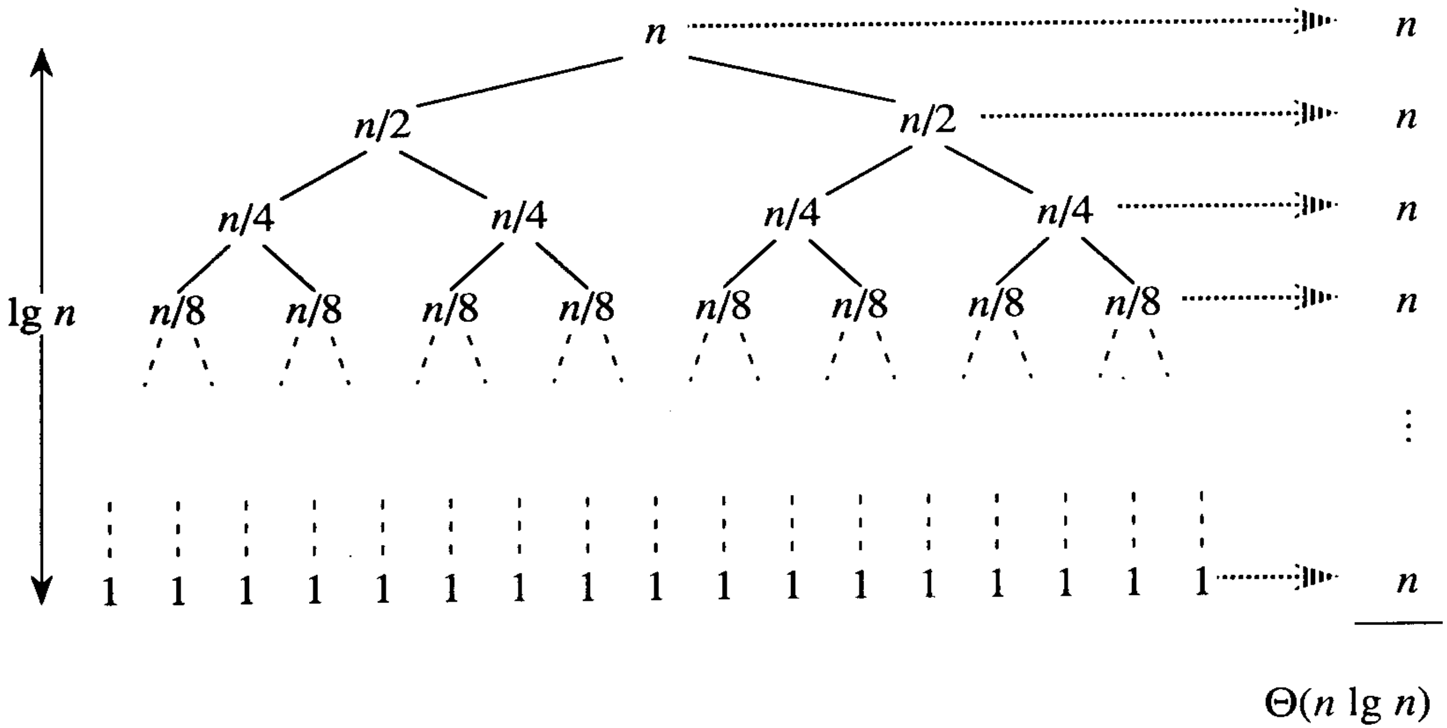
Recursion: recursively sort the lists L and R.



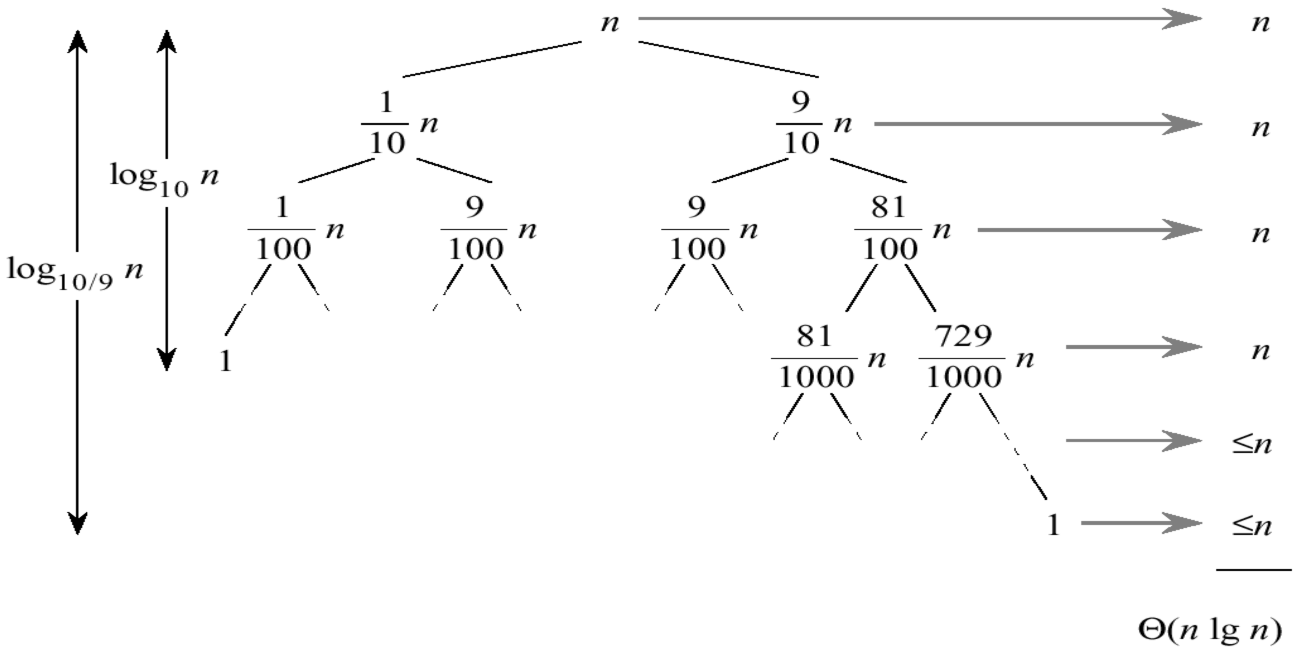
Worst Case Partitioning of Quick Sort



Best Case Partitioning of Quick Sort



Average Case of Quick Sort



Randomized Quick Sort

<u>Randomized-Partition(*A, p, r*)</u>

1. $i \leftarrow Random(p, r)$
2. exchange $A[r] \leftrightarrow A[i]$
3. **return Partition**(A, p, r)

<u>Randomized-Quicksort(*A, p, r*)</u>

1. **if** $p < r$
2. **then** $q \leftarrow$ **Randomized-Partition**(A, p, r)
3. **Randomized-Quicksort**$(A, p, q-1)$
4. **Randomized-Quicksort**$(A, q+1, r)$

Randomized Quick Sort

- Exchange $A[r]$ with an element chosen at random from $A[p \ldots r]$ in **Partition**.

- The pivot element is equally likely to be any of input elements.

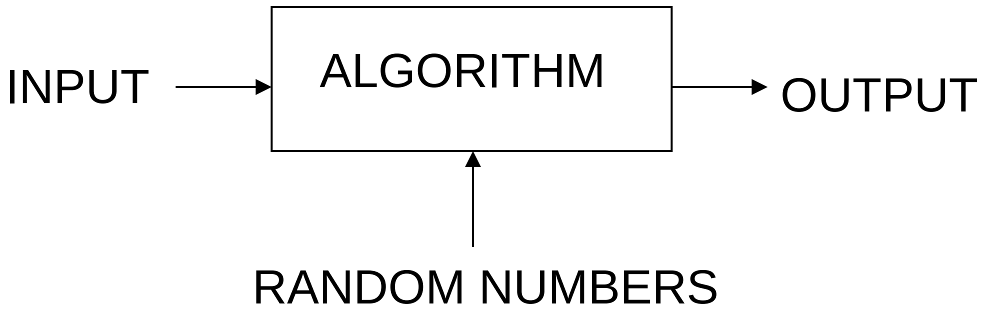
- *For any given input, the behavior of Randomized Quick Sort* is determined not only by the input but also by the *random choices of the pivot.*

- We add randomization to Quick Sort to obtain for any input the expected performance of the algorithm to be good.

Deterministic Algorithms

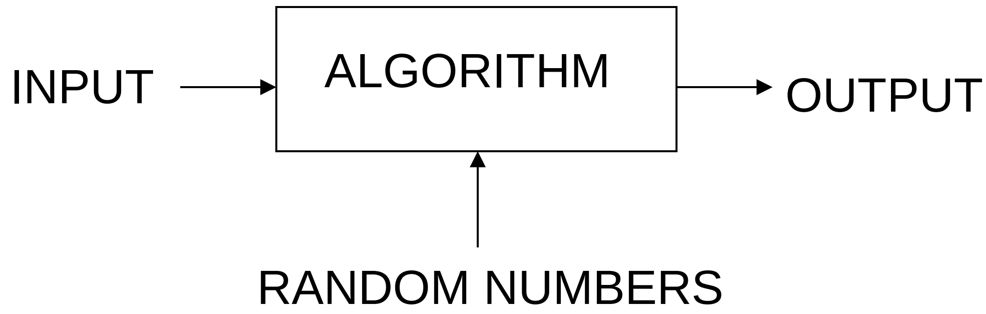
Goal: Prove for all input instances the algorithm solves the problem correctly and the number of steps is bounded by a polynomial in the size of the input.

Randomized Algorithms



- In addition to input, algorithm takes a source of random numbers and makes random choices during execution;
- Behavior can vary even on a fixed input;

Las Vegas Randomized Algorithms



Goal: Prove that for all input instances the algorithm solves the problem correctly and the expected number of steps is bounded by a polynomial in the input size.

Note: The expectation is over the random choices made by the algorithm.

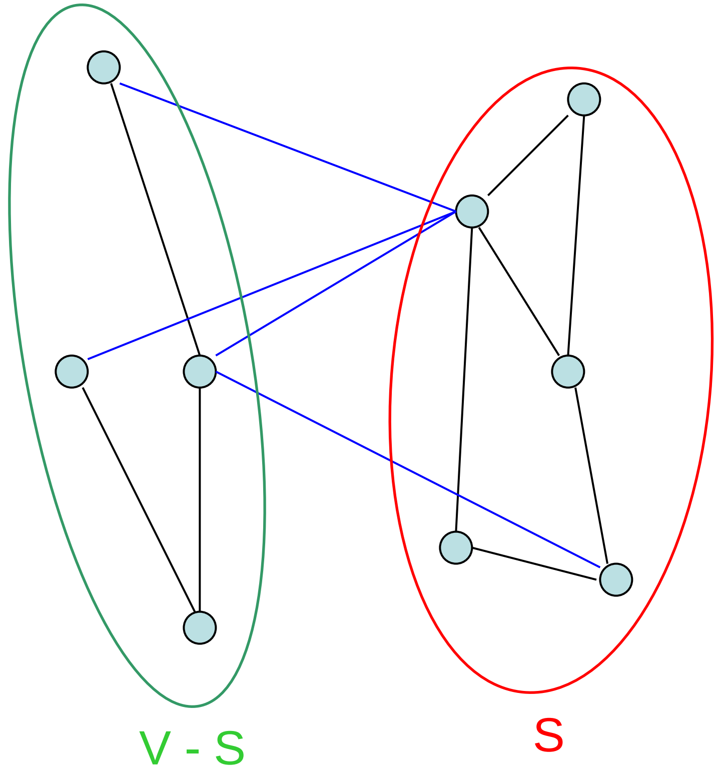
Probabilistic Analysis of Algorithms

Input is assumed to be from a probability distribution.

Goal: Show that for all inputs the algorithm works correctly and for most inputs the number of steps is bounded by a polynomial in the size of the input.

Min-cut for Undirected Graphs

Given an undirected graph, a global *min-cut* is a cut (S,V-S) minimizing the number of *crossing edges*, where a crossing edge is an edge (u,v) s.t. u∈S and v∈ V-S.



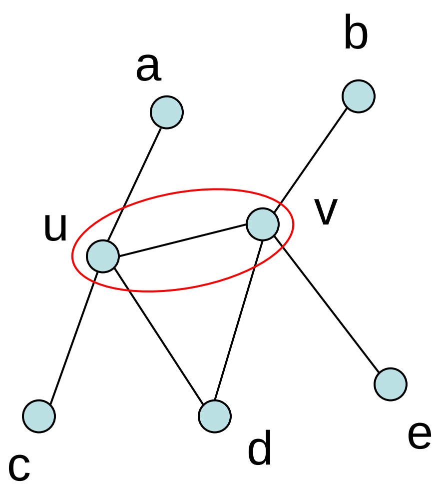
Graph Contraction

For an undirected graph G, we can construct a new graph G' by *contracting* two vertices u, v in G as follows:

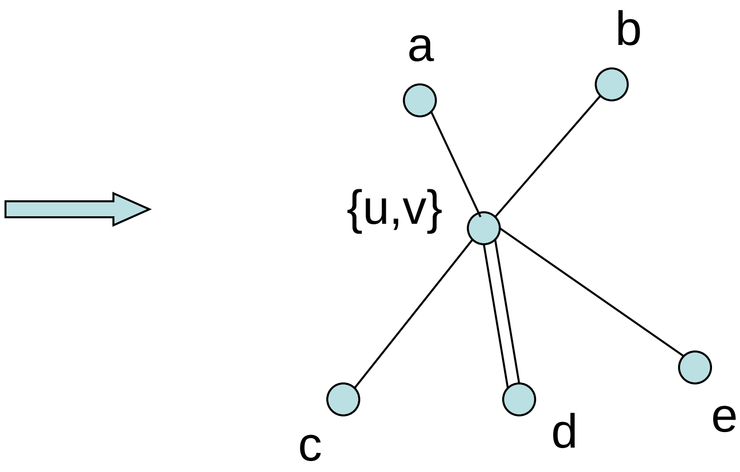
– u and v become one vertex {u,v} and the edge (u,v) is removed;

– the other edges incident to u or v in G are now incident on the new vertex {u,v} in G';

Note: There may be multi-edges between two vertices. We just keep them.

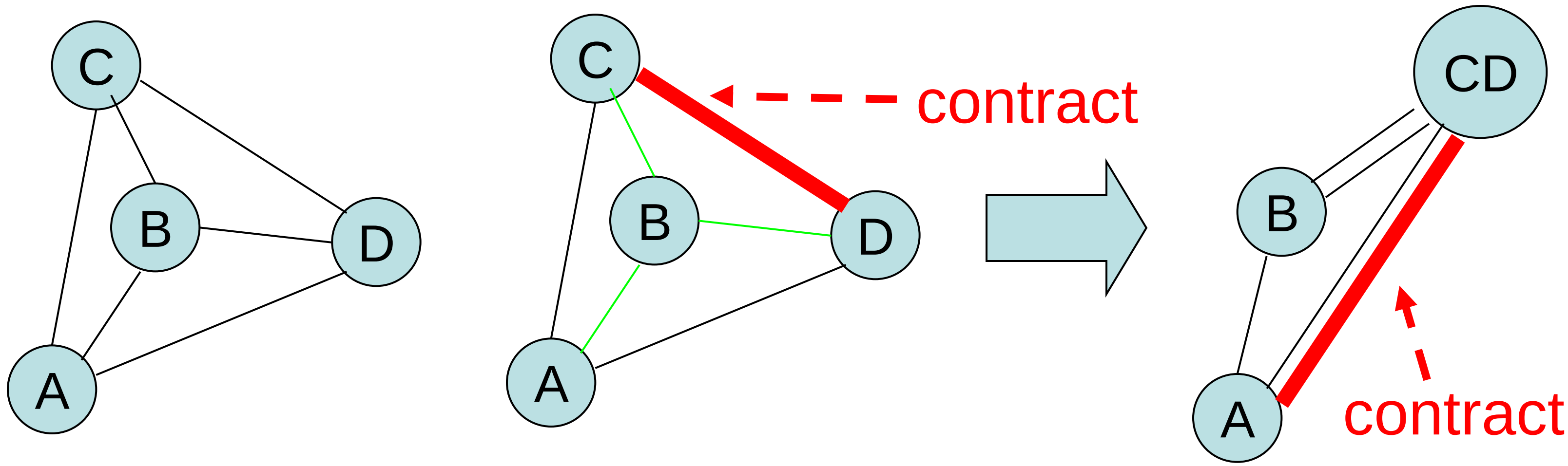


Graph G

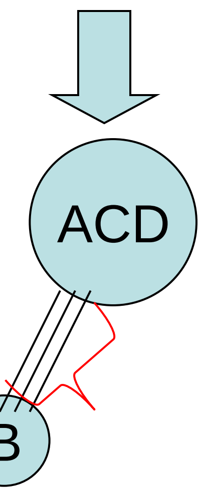


Graph G'

Karger's Min-cut Algorithm



(i) Graph G (ii) Contract nodes C and D (iii) contract nodes A and CD



Note: C is a cut but not necessarily a min-cut.

(Iv) Cut C=*{(A,B), (B,C), (B,D)}*

Karger's Min-cut Algorithm

For i = 1 to $100n^2$

    repeat

        randomly pick an edge (u,v)

        contract u and v

    until two vertices are left

    $c_i \leftarrow$ the number of edges between them

Output **mini c_i**

Key Idea

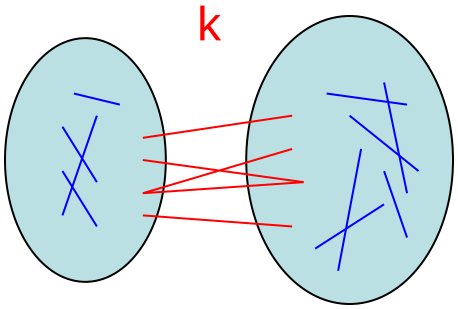
- Let $\underline{C^* = \{c_1^*, c_2^*, \ldots, c_k^*\}}$ be a min-cut in G and C^i be a cut determined by Karger's algorithm during some iteration i.

- C^i will be a min-cut for G if during iteration "i" none of the edges in C* are contracted.

- If we can show that with *prob. $\Omega(1/n^2)$, where $n = |V|$, C^i will be a min-cut*, then by *repeatedly obtaining min-cuts $O(n^2)$ times* and taking minimum gives the min-cut with high prob.

Analysis of Karger’s Min-Cut Algorithm

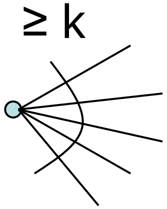
Analysis of Karger's Algorithm



Let k be the number of edges of min cut (S, V-S).

If we never picked a crossing edge in the algorithm, then the number of edges between two last vertices is the correct answer.

The probability that in step 1 of an iteration a crossing edge is not picked = (|E|-k)/|E|.



By def of min cut, we know that each vertex v has degree at least k, Otherwise the cut ({v}, V-{v}) is lighter.

Thus $|E| \geq nk/2$ and $(|E|-k)/|E| = 1 - k/|E| \geq 1-2/n$.

Analysis of Karger's Algorithm

- In step 1, Pr [no crossing edge picked] >= 1 – 2/n
- Similarly, in step 2, Pr [no crossing edge picked] ≥ 1-2/(n-1)
- In general, in step j, Pr [no crossing edge picked] ≥ 1-2/(n-j+1)
- Pr {the n-2 contractions never contract a crossing edge}
  - = Pr [first step good]

    * Pr [second step good after surviving first step]

    * Pr [third step good after surviving first two steps]

    * …

    * Pr [(n-2)-th step good after surviving first n-3 steps]

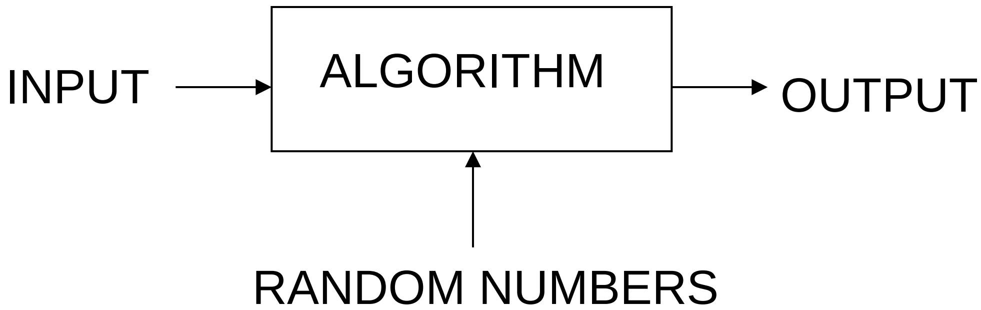
    ≥ (1-2/n) (1-2/(n-1)) … (1-2/3)

    = [(n-2)/n] [(n-3)(n-1)] … [1/3] = 2/[n(n-1)] = $\Omega(1/n^2)$

Introduction to Randomized Algorithms:
Monte Carlo Randomized Algorithm

This part of slides come from Srikrishnan Divakaran DA-IICT

Monte Carlo Randomized Algorithms



Goal: Prove that the algorithm

- with high probability solves the problem correctly;
- for every input the expected number of steps is bounded by a polynomial in the input size.

Note: The expectation is over the random choices made by the algorithm.

Monte Carlo versus Las Vegas

- A Monte Carlo algorithm runs produces an answer that is correct with non-zero probability, whereas a Las Vegas algorithm always produces the correct answer.

- The running time of both types of randomized algorithms is a random variable whose expectation is bounded say by a polynomial in terms of input size.

- These expectations are only over the random choices made by the algorithm independent of the input. Thus independent repetitions of Monte Carlo algorithms drive down the failure probability exponentially.

Motivation for Randomized Algorithms

- Simplicity;
- Performance;
- Reflects reality better (Online Algorithms);
- For many hard problems helps obtain better complexity bounds when compared to deterministic approaches;

Analysis of Randomized Quick Sort

Linearity of Expectation

If X_1, X_2, …, X_n are random variables, then

$$E\left[\sum_{i=1}^{n} Xi\right] = \sum_{i=1}^{n} E[Xi]$$

Notation

z_2	z_9	z_8	z_3	z_5	z_4	z_1	z_6	z_{10}	z_7
2	9	8	3	5	4	1	6	10	7

- Rename the elements of A as $z_1, z_2, \ldots, z_n$, with z_i being the i^{th} smallest element (Rank "i").
- Define the set $Z_{ij} = \{z_i, z_{i+1}, \ldots, z_j\}$ be the set of elements between z_i and z_j, inclusive.

Expected Number of Total Comparisons in PARTITION

Let X_{ij} = I {z_i is compared to z_j } ← indicator random variable

Let X be the total number of comparisons performed by the algorithm. Then

$$\left[X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \right]$$

The expected number of comparisons performed by the algorithm is

$$E[X] = E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left[X_{ij}\right]$$

by linearity of expectation

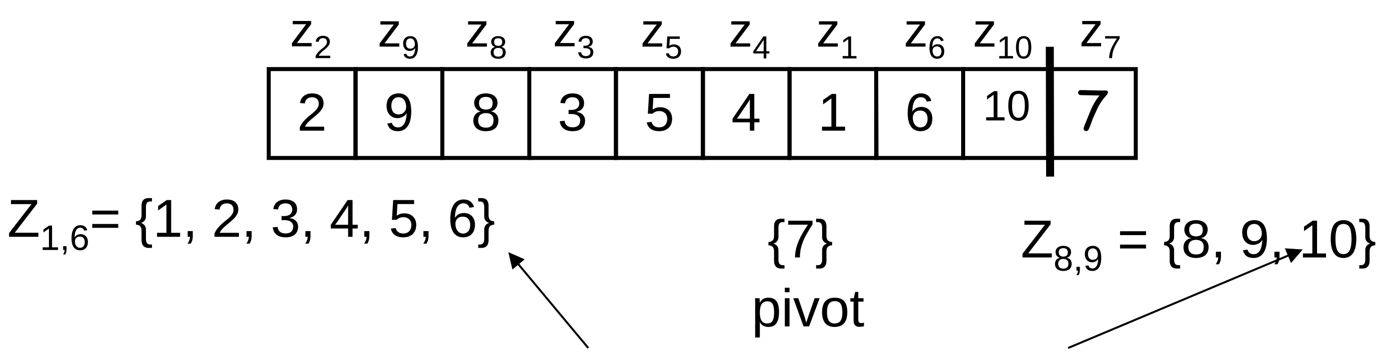
$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr\{z_i \textit{ is compared to } z_j\}$$

Comparisons in PARTITION

<u>Observation 1</u>: Each pair of elements is compared **at most once** during the entire execution of the algorithm

- Elements are compared only to the pivot point!
- Pivot point is excluded from future calls to PARTITION

<u>Observation 2</u>: Only the pivot is compared with elements in both partitions



Elements between different partitions are <u>never</u> compared

Comparisons in PARTITION

z_2	z_9	z_8	z_3	z_5	z_4	z_1	z_6	z_{10}	z_7
2	9	8	3	5	4	1	6	**10**	7

$Z_{1,6}$= {1, 2, 3, 4, 5, 6} {7} $Z_{8,9}$ = {8, 9, 10}

$$\Pr\{z_i \textit{ is compared to } z_j\}?$$

<u>Case 1</u>: pivot chosen such as: $z_i < x < z_j$

- z_i and z_j will never be compared

<u>Case 2</u>: z_i or z_j is the pivot

- z_i and z_j will be compared
- only if one of them is chosen as pivot before any other element in range z_i to z_j

Expected Number of Comparisons in PARTITION

Pr {Z_i is compared with Z_j}

= Pr{Z_i or Z_j is chosen as pivot before other elements in $Z_{i,j}$} = 2 / (j-i+1)

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr\{z_i \text{ is compared to } z_j\}$$

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} < \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} = \sum_{i=1}^{n-1} O(\lg n)$$

$$= O(n\lg n)$$

Basic Analytical Tools

Tail Bounds

- In the analysis of randomized algorithms, we need to know how much does an algorithms run-time/cost deviate from its expected run-time/cost.
- That is we need to find an upper bound on Pr[X deviates from E[X] a lot]. This we refer to as the tail bound on X.

Markov and Chebyshev's Inequality

Markov's Inequality If $X \geq 0$, then

$$\Pr[X \geq a] \leq E[X]/a.$$

Proof. Suppose $\Pr[X \geq a] > E[X]/a$. Then

$$E[X] \geq a \cdot \Pr[X \geq a] > a \cdot E[X]/a = E[X].$$

Chebyshev's Inequality: $\Pr[\ |X-E[X]| \geq a\ ] \leq Var[X] / a^2$.

Proof.

$$\begin{aligned}
&\Pr[\ |X-E[X]| \geq a\ ] \\
=\ & \Pr[\ |X-E[X]|^2 \geq a^2\ ] \\
=\ & \Pr[\ (X-E[X])^2 \geq a^2\ ] \\
\leq\ & E[(X-E[X])^2] / a^2 \qquad \text{// Markov on } (X-E[X])^2 \\
=\ & Var[X] / a^2
\end{aligned}$$