Basics on Probability

Josep Díaz Maria J. Serna Conrado Martínez
U. Politècnica de Catalunya

RA-MIRI 2022–2023

- Overview on basic probability
- The principle of deferred decisions
- Checking matrix multiplication
- The minimum cut problem

Review of basic mathematics

- Arithmetic Series: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2} = \Theta(n^2)$.
- Geometric Series: for $x \neq 1$, $\sum_{i=0}^{n} x^i = \frac{x^{n+1}-1}{x-1}$.
- Geometric Series: for $|x| < 1$, $\sum_{i=0}^{n} x^i = \frac{1}{1-x}$.
- Harmonic Series: for n finite,

$$H_n = \sum_{i=1}^{n} \frac{1}{i} = \ln n + \mathcal{O}(1).$$

Note that if $n \to \infty$ then $\sum_{i=1}^{n} \frac{1}{i}$ diverges.

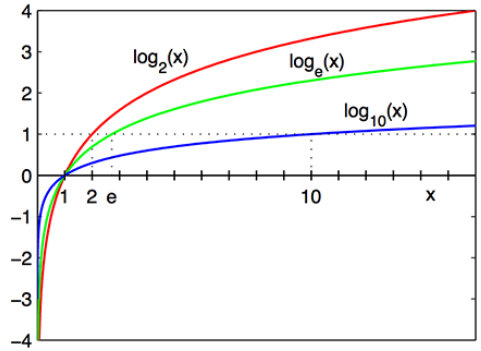
Review of basic mathematics: Log and Exponential

$\log_b n = x$ means $n = b^x$,

$\log(xy) = \log x + \log y$

$\log\left(x^{f(x)}\right) = f(x)\log x \Rightarrow 2^{\lg n} = n.$

$\log_a x = \frac{\log_b x}{\log_a b}$



Recall:

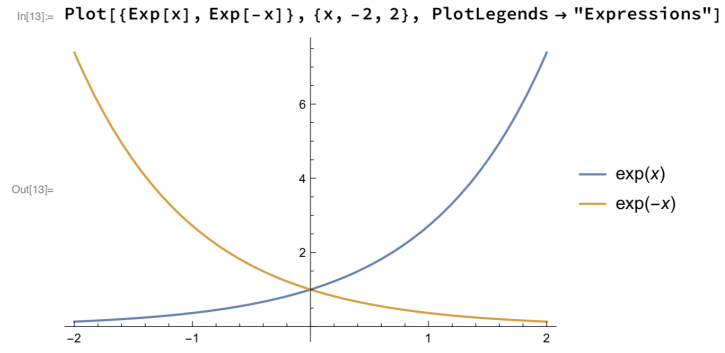
$$\frac{d}{dx}\ln(f(x)) = \frac{\frac{d(f(x))}{dx}}{f(x)}.$$

$$\frac{d}{dx}\ln x = \frac{1}{x}.$$

$\lg = \log_2$, $\ln = \log_e$, $\log = \log_{10}$

Review of basic mathematics: Exponential

$\ln n = \log_e n = x$ means $n = e^x$,
where
$e = \lim_{n\to\infty}(1 + \frac{1}{n})^n \sim 2.71...$
$e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$.
$e^{-x} = \lim_{n\to\infty}(1 - \frac{x}{n})^n$.
$\frac{d}{dx}e^x = e^x$.



Recall Taylor: $f(x)$ differentiable at a: $f(x) = \sum_{i=0}^{\infty} \frac{f^i(a)}{i!}(x-a)^i$

Therefore $e^x = 1 + x + \frac{x^2}{2!} + \cdots$, $e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$

$\boxed{e^x > 1 + x}$ and $\boxed{e^{-x} > 1 - x}$

In fact, when x is very small $0 < x \ll 1$: $\boxed{e^{-x} \sim 1 - x}$

$e^{-0.4} = 0.67032$, $e^{-0.1} = 0.904837$, $e^{-0.01} = 0.99005$

Binomial

- Stirling: $n! = \sqrt{2\pi n}(\frac{n}{e})^n + \gamma + \mathcal{O}(1/n)$,
- Binomial coefficients: $\binom{n}{k} = \frac{n!}{(n-k)!k!}$
- Binomial Thm.: $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$.
  $\therefore (1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i = 1 + nx + \frac{n(n-1)}{2} x^2 + \cdots + x^n$
- Important: $\boxed{\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} \leqslant \left(\frac{ne}{k}\right)^k}$
- Also useful: if $k = o(\sqrt{n})$ then $\binom{n}{k} \sim \frac{n^k}{k!}$

Why using asymptotic notation?

Considering that an instance with size $n = 1$ takes 1 μ second:

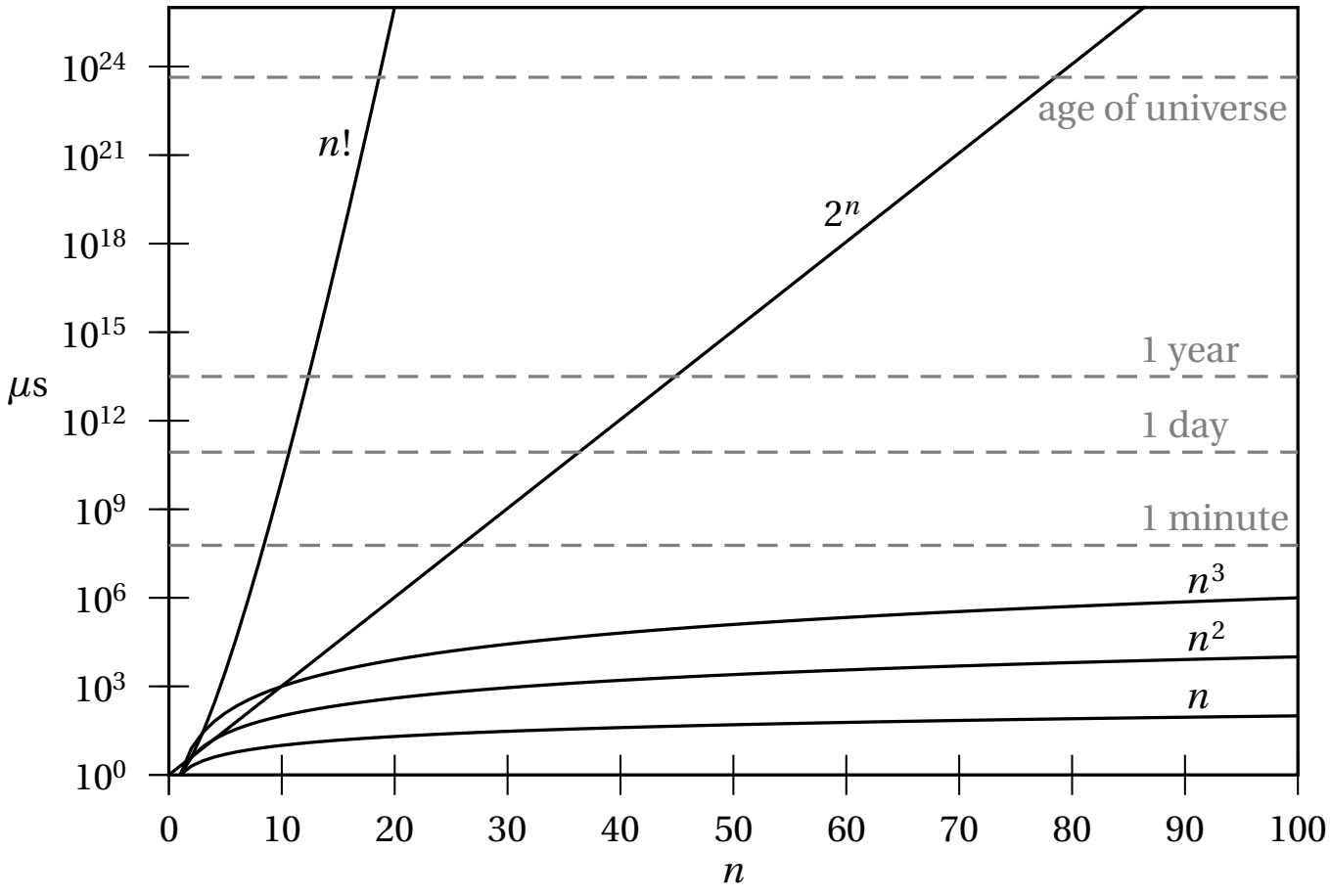


Table of computing times according to the size of an instance.

Recall: Asymptotic notation

Symbol	$L = \lim_{n\to\infty} \frac{f(n)}{g(n)}$	Relational notation
$f(n) = \mathcal{O}(g(n))$	$L < \infty$	$f \preccurlyeq g$
$f(n) = \Omega(g(n))$	$L > 0$	$f \succcurlyeq g$
$f(n) = \Theta(g(n))$	$0 < L < \infty$	$f \asymp g$
$f(n) = o(g(n))$	$L = 0$	$f \prec g$
$f(n) = \omega(g(n))$	$L = \infty$	$f \succ g$
$f(n) = g(n) + o(g(n))$	$L = 1$	$f(n) \sim g(n)$

For ex. $\log_a x = \Theta(\log_b x)$, for any $a, b > 0$.

Remember: Basic Combinatorics

For a set S with n elements

- The permutations of S are all the ordered sequences of length n without repetition.
  Ex.: $S = \{a, b, c\}$ then $abc, acb, bac, bca, cab, cba$.
  There are $n!$ permutations of S.
- The k-permutations of S ($k \leqslant n$) are all the ordered sequences of length k without repetition.
  The 2-permutations of $\{a, b, c\}$ are ab, ac, ba, bc, ca, cb.
  There are $P(n, k) = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1) = n^{\underline{k}}$ k-permutations of S.
- For $m > n$ the number of ordered m-sequences with repetitions that we can form with elements in S is n^m.
  Ex. The number of binary sequences with length 5 is 2^5.

k-Combination: Binomial

A k-combination of S with ($k \leqslant n$) are all the non-ordered sequences of length k without repetition. Ex.:
$S = \{a, b, c\}, k = 2$ then we get ab, ac, bc

This is the same as the number of different k-subsets, i.e.,

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Notice $\binom{n}{0} = \binom{n}{n} = 1$ and $\binom{n}{k} = \binom{n}{n-k}$.

Experiments and Events

Probability space (Ω): the set of outcomes associated with an experiment.

Basic events: the elements in Ω.
Event: $\mathrm{E} \subseteq \Omega$, i.e. an event is any collection of outcomes.

Example: Flip two coins:

- Basic events $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$. $|\Omega| = 4$.
- Non-basic event: Let A be the event of having at least one H, then $A = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\}$.

Given Ω, define $\mathcal{F}$ is the set of all events in the powerset of Ω, $\mathcal{F} = \mathcal{P}(\Omega)$.

For any event $\mathrm{E} \in \mathcal{F}$, let $\bar{\mathrm{E}} \in \mathcal{F}$ the set of events $\bar{\mathrm{E}} = \Omega \setminus \mathrm{E}$; $\bar{\mathrm{E}}$ is the complementary of the event E

Probability

Given the set of events $\mathcal{F}$ in Ω, a probability function (distribution) $\mathbb{P}[\cdot] : \mathcal{F} \to [0, 1]$ is a function such that:

1. For any event $A \in \mathcal{F}$: $0 \leqslant \mathbb{P}[A] \leqslant 1$, $\mathbb{P}[\Omega] = 1$, $\mathbb{P}[\emptyset] = 0$.
2. Given all basic events $\{E_i\}_{i=1}^n$, $\sum_{i=1}^n \mathbb{P}[E_i] = 1$,
3. If $\{A_j\}_{j=1}^k$ are mutually exclusive events then

$$\mathbb{P}\left[\cup_{j=1}^k A_j\right] = \sum_{j=1}^k \mathbb{P}\left[A_j\right].$$

In a probability space $(\Omega, \mathcal{F}, \mathbb{P}[\cdot])$, the set of basic events $\{E_i\}_{i=1}^n$ forms a partition of Ω, i.e., they are mutually disjoint, therefore $\sum_{i=1}^n \mathbb{P}[E_i] = 1$ follows from 1 and 3.

Uniform distribution

In a finite discrete probability space, $|\Omega| = n$, the uniform distribution assigns to any basic event E_i identical probability: $\mathbb{P}[E_i] = \frac{1}{n}$.

Given a probability space we select uniformly at random (u.a.r.) an element in Ω if we choose with equal probability among all basic events.

Examples:

Flip 3 coins: $|\Omega| = 2^3 = 8$, so probability of choosing u.a.r. : $\mathbb{P}[000] = \mathbb{P}[011] = 1/8$.

If A is the event that we choose an element with two 1's, $\mathbb{P}[A] = \mathbb{P}[011] + \mathbb{P}[101] + \mathbb{P}[110] = 3/8$

More on events

In general, an event A is a collection of outcomes, i.e., $A \subseteq \Omega$

Given an event $A \subseteq \Omega$ we define its probability:

$$\mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\omega],$$

Example

Flip a fair coin. If it comes up heads, roll a 3-sided die; if it comes up tails, roll a 4-sided die. What is the probability that the die roll is at least 3?

$\Omega = \{(H,1),(H,2),(H,3),(T,1),(T,2),(T,3),(T,4)\}$, $|\Omega| = 7$

As $A = \{(H,3),(T,3),(T,4)\}$

$$\mathbb{P}[A] = \mathbb{P}[(H,3)] + \mathbb{P}[(T,3)] + \mathbb{P}[(T,4)]$$

$$= \frac{1}{2} \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} \cdot \frac{1}{4} = 1/6 + 1/4 = 5/12$$

Examples

Example

We have a unit square $\mathcal{S}$ with side 1, and inside a circle $\mathcal{C}$ centered at the central point of $\mathcal{S}$ and of radius $r = 1/4$. If we throw u.a.r. a point to $\mathcal{S}$, which is the probability it hits inside $\mathcal{C}$?

The probability is $= \frac{\text{Area } \mathcal{C}}{\text{Area } \mathcal{S}} = \pi(1/4)^2 = 0.1965$

Example

A bag contains 100 balls, 50 red and 50 blue. We select 5 balls independently and u.a.r. What is the probability that 3 are blue and 2 are red?
The total number of outcomes $|\Omega| = \binom{100}{5}$. Therefore the probability is:

$$\frac{\binom{50}{3}\binom{50}{2}}{\binom{100}{5}} = \frac{6125}{19206} \approx 0.318910757\ldots$$

Some consequences of the probability properties

Given $A, B, C \in \mathcal{F}$:

- $\mathbb{P}\left[\bar{A}\right] = 1 - \mathbb{P}[A]$.
- If $A \subseteq B$ then $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \setminus A] \geqslant \mathbb{P}[A]$.
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

  Pf. Events $(A \setminus B)$, $(B \setminus A)$ and $(A \cap B)$ are disjoint.
- **Inclusion-Exclusion: 3 events**

$$\begin{aligned}\mathbb{P}[A \cup B \cup C] = \mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] \\ - \mathbb{P}[A \cap B] - \mathbb{P}[B \cap C] - \mathbb{P}[A \cap C] \\ + \mathbb{P}[A \cap B \cap C]\,.\end{aligned}$$

Inclusion-Exclusion and Union-Bound

Inclusion-Exclusion: General

Given n events $\{A_1, \ldots, A_n\}$,

$$\mathbb{P}[\cup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}\left[A_i \cap A_j\right] + \sum_{i<j<k} \mathbb{P}\left[A_i \cap A_j \cap A_k\right] + \cdots (-1)^{l+1} \sum_{i_1<\cdots<i_l} \mathbb{P}\left[\cap_{r=1}^{l} A_{i_r}\right] + \cdots$$

Very useful upper-bound to the probability of non-exclusive events:Union-Bound. Given non-independent events $\{A_i\}_{i=1}^{n}$,

$$\mathbb{P}[\cup_{i=1}^{n} A_i] \leqslant \sum_{i=1}^{n} \mathbb{P}[A_i]\,.$$

Independent and correlated events

Given events A, B in Ω, we say they are independent (mutually independent) if $\mathbb{P}[A \cap B] = \mathbb{P}[A] \times \mathbb{P}[B]$, otherwise they are said to be correlated or dependent.
Events $A_1, A_2, \cdots A_n$ are independent if

$$\mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n] = \prod_{i=1}^{n} \mathbb{P}[A_i] .$$

Notice the basic events in Ω are not independent, although they are disjoint.
For example, if we flip a coin, and E_1 is the event of (H), and E_2 is the event of (T), then $\mathbb{P}[E_1]\,\mathbb{P}[E_2] = \frac{1}{4} \neq 0 = \mathbb{P}[E_1 \cap E_2]$

But if the experiment is flipping twice a coin and E_1 is the event of (H) in the 1st flip and $E_2 =$ event of (H) in the 2nd flip, then E_1 and E_2 are independent.

Independent and correlated events

Toss 2 fair coins and consider the events: A, there is at least 1 head, and B, there is at least one tail.

$\Omega = \{\mathrm{HH}, \mathrm{TT}, \mathrm{TH}, \mathrm{HT}\} \Rightarrow \mathbb{P}[\mathrm{A}] = \frac{3}{4} = \mathbb{P}[\mathrm{B}] = \frac{3}{4}$

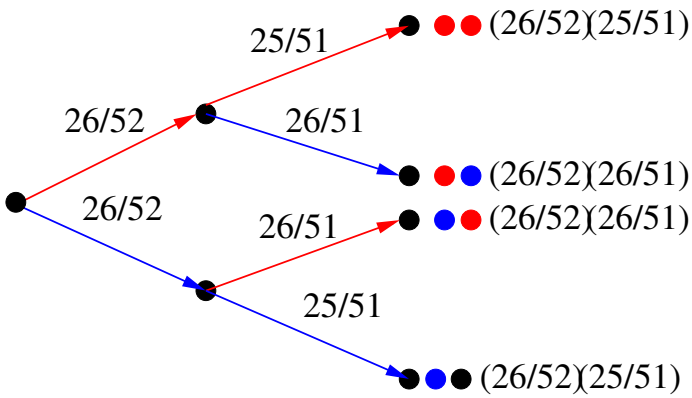
but $\mathbb{P}[\mathrm{A} \cap \mathrm{B}] = \frac{2}{4} \neq \frac{3}{4}\frac{3}{4} = \frac{9}{16}$

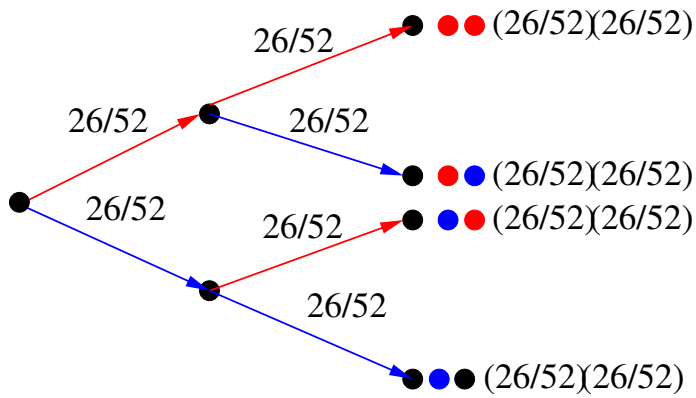
Therefore A and B, are dependent (correlated).

Sampling with replacement simplifies life

Important Example: We draw sequentially 2 cards from a deck with 52 cards, where 26 of the cards are red and the other half blue. Let R_1 be the event of drawing a red card on the first trial and R_2 the event of drawing a red card on the second trial.

If the draws are with replacement R_1 and R_2 are independent, if they are without replacement R_1 and R_2 are not independent.





Without replacement: $\mathbb{P}[R_1 \cap R_2] = \frac{26}{52} \cdot \frac{25}{51} \neq \mathbb{P}[R_1] \cdot \mathbb{P}[R_2]$

With replacement: $\mathbb{P}[R_1 \cap R_2] = \frac{26}{52} \cdot \frac{26}{52} = \mathbb{P}[R_1] \cdot \mathbb{P}[R_2]$

Formal proof sampling without replacement are not independent events

Draw sequentially 2 cards from a 52 deck. Let R_1 be the event of drawing a red card on the first trial and R_2 the event of drawing a red card on the second trial. If we draw without replacement, R_1 and R_2 are not independent.

Let B_1 event of drawing a black card 1st. trial.

Recall: $\mathbb{P}[\mathrm{R}_1] = \frac{26}{52}$ and $\mathbb{P}[\mathrm{B}_1] = \frac{26}{52}$.

Need $\mathbb{P}[\mathrm{R}_1 \cap \mathrm{R}_2] =^{?} \mathbb{P}[\mathrm{R}_1]\,\mathbb{P}[\mathrm{R}_2]$

After R_1, prob. drawing another $\mathrm{R} = \frac{25}{51} \Rightarrow \mathbb{P}[\mathrm{R}_1 \cap \mathrm{R}_2] = \frac{26}{52}\frac{25}{51}$

So $\mathbb{P}[\mathrm{R} \text{ then } \mathrm{R}] = \frac{26}{52}\frac{25}{51}$ and $\mathbb{P}[\mathrm{B} \text{ then } \mathrm{R}] = \frac{26}{52}\frac{26}{51}$

$\Rightarrow \mathbb{P}[\mathrm{R}_2] = \frac{26}{52}\frac{25}{51} + \frac{26}{52}\frac{26}{51} = \frac{26}{52}$.

$$\therefore \mathbb{P}[\mathrm{R}_1 \cap \mathrm{R}_2] = \frac{26}{52}\frac{25}{51} \neq \frac{26}{52}\frac{26}{52} = \mathbb{P}[\mathrm{R}_1]\,\mathbb{P}[\mathrm{R}_2]\,.$$

Conditional probability

One of the important concepts in probability is conditioning, which means revising probabilities on an event A based on *partial information* that we know, i.e. based in another event B.

Flip 2 fair coins. Given that event B that one of them is H, what is the probability of the event A that both of them are H?
$\mathbb{P}[A|B] = 1/3$, *as the information* B *reduces the probability space to* $\{TH, HT, HH\}$*, each one with probability* $1/3$*.*

Formal definition of conditional probability:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B \cap A]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A]\,\mathbb{P}[A]}{\mathbb{P}[B]}.$$

In previous ex.: $\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{1/4}{3/4}$.

Alternative definition of independence:
A and B are independent iff $\mathbb{P}[A|B] = \mathbb{P}[A]$.

The Russian roulette

Two people play one round of Russian roulette. The gun is a revolver with six chambers, all empty. The players put two bullets into adjacent chambers of the barrel. The first player takes the gun and spins the barrel, then he puts the gun in his head and pulls the trigger and no bullet!
He gives the gun to the second player. Which would be better for the second player, to spin the barrel first, or just pull the trigger?

The Russian roulette

If player 2 spins the barrel, the probability of getting a bullet is $2/6 = 1/3$ so the probability of survival is $1 - 1/3 = 2/3$, i.e 66.66%

If he does not spin the barrel, we are conditioning to the fact that we are positioned right after one of the 4 empty chambers. Only one of the empty chambers leads to one with a bullet. So the probability of having a bullet is $1/4$, therefore the probability of non-having a bullet is $3/4 = 75\%$. So it is better no to spin the barrel.

Total probability law

When dealing with conditional probability, it seems that first we have to compute the probabilities involved in a random experiment, and then we can calculate the conditional probabilities.
In practice we use conditional probabilities to *reduce* the calculation of probabilities for events.

Total Probability Law If a set of events $\{E_i\}_{i=1}^n$ is a **partition** of Ω and $A \in \mathcal{F}$ is a event, then

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \cap E_i] = \sum_{i=1}^{n} \mathbb{P}[A|E_i]\,\mathbb{P}[E_i]\,.$$

Principle of deferred decisions

Not to assume that the entire set of random choices is made in advance. Rather, at each step of the process concentrate only on the random choices that are relevant to the algorithm outcome

When applicable it provides a simplified probability space to perform the probabilistic analysis.

Analyzing the Clock Solitaire game

From MR 3.5

The Clock Solitaire game: randomly shuffle a standard pack of 52 cards. Then, split the cards into 13 piles of 4 cards each; label piles as A, 2, . . . , 10, J, Q, K; take the first card from the "K" pile; take the next card from the pile "X", where X is the value of the previous card taken; repeat until:

- either all cards removed ("win")
- or you get stuck ("lose")

We want to evaluate the probability of "win".

Analyzing the Clock Solitaire game

Game termination?
The last card we take before the game ends (either winning or loosing) is a “K”.

Let us assume that at iteration j we draw card X but the pile X is empty (thus the game terminates).

Suppose $X \neq K$. Because pile X is empty and $X \neq K$, we must have already drawn (prior to draw j) the 4 cards numbered X. But then, we can not draw an X card at the jth iteration, a contradiction.

There is no contradiction if the last card is a “K” and all other cards have been already removed (in that case the game terminates with win).

Analyzing the Clock Solitaire game

Game win?

We win if and only if the fourth “K” card is drawn at the 52 iteration.

Whenever we draw for the 1st, 2nd or 3rd time a “K” card, the game does not terminate because the “K” pile is not empty so we can continue.

When the fourth K is drawn at the 52nd iteration then all cards are removed and the game’s result is “win”

Analyzing the Clock Solitaire game

The probability of win?
According to the previous observations

$$\begin{aligned}\mathbb{P}[\text{win}] &= \mathbb{P}[\text{4th “K” at the 52nd iteration}] \\ &= \frac{\text{\#game evolutions: 52nd card = 4th “K”}}{\text{\#all game evolutions}}\end{aligned}$$

Considering all possible game evolutions is a rather naive approach since we have to count all ways to partition the 52 cards into 13 distinct piles, with an ordering on the 4 cards in each pile. This complicates the probability evaluation because of the dependence introduced by each random draw of a card.

We define another probability space that better captures the random dynamics of the game evolution.

Analyzing the Clock Solitaire game

The principle of deferred decisions
Basic idea: rather than fix (and enumerate) the entire set of potential random choices in advance, instead let the random choices unfold with the progress of the random experiment.

In this particular game at each draw any card not drawn yet is equally likely to be drawn.

A winning game corresponds to a dynamics where the first 51 random draws include 3 “K” cards exactly.

This is equivalent to draw the 4th “K” at the 52nd iteration.
So we “forget” how the first 51 draws came out and focus on the 52nd draw, which must be a “K”.

Analyzing the Clock Solitaire game

The probability of win:
We actually have 13x4=52 distinct positions (13 piles, 4 positions each) where 52 distinct cards are placed. This gives a total of 52! different placements.

Each game evolution actually corresponds to an ordered permutation of the 52 cards.

The winning permutations are those where the 52nd card is a "K" (4 ways) and the 51 preceding cards are arbitrarily chosen (51!). Thus:

$$\mathbb{P}[\text{win}] = \frac{4\,51!}{52!} = \frac{4}{52} = \frac{1}{13}.$$

A simpler way to get the same: The probability is $\frac{1}{13}$ because of symmetry (e.g., the type of the 52nd card is random uniform among all 13 types).

The idea was to defer, i.e., first consider the last choice and then conditionally the previous ones!

Analyzing the Clock Solitaire game

The probability of win:
We actually have 13x4=52 distinct positions (13 piles, 4 positions each) where 52 distinct cards are placed. This gives a total of 52! different placements.

Each game evolution actually corresponds to an ordered permutation of the 52 cards.

The winning permutations are those where the 52nd card is a "K" (4 ways) and the 51 preceding cards are arbitrarily chosen (51!). Thus:

$$\mathbb{P}[\text{win}] = \frac{4\,51!}{52!} = \frac{4}{52} = \frac{1}{13}.$$

A simpler way to get the same: The probability is $\frac{1}{13}$ because of symmetry (e.g., the type of the 52nd card is random uniform among all 13 types).

The idea was to defer, i.e., first consider the last choice and then conditionally the previous ones!

Checking matrix multiplication

Problem: Given 3 square matrices ($n \times n$), A, B and C, we want to see if $A \times B = C$.
Easy solution: compute $A \times B$ and compare with C.

$n \times n$ matrix multiplication:

1. Naive algorithm: $\mathcal{O}(n^3)$
2. Strassen (1969): $\mathcal{O}(n^{2.81})$
3. Coppersmith-Winograd (1987): $\mathcal{O}(n^{2.376})$
4. Vassilevska (2015): $\mathcal{O}(n^{2.373})$

Can we (randomly) check in $\mathcal{O}(n^2)$ if $A \times B = C$?

Freivald's algorithm

From MU 1.3, MR 3.5 Freivald's algorithm (1977) checks if $A \times B = C$ for three given $n \times n$ matrices A ,B, and C

```
procedure FREIVALD(A, B, C)
    Choose u.a.r. r⃗ ∈ {0, 1}^n
    if A(Br⃗) = Cr⃗ then
        return true
    else
        return false
    end if
end procedure
```

Choosing u.a.r. $\vec{r}$ can be done choosing independently with probability $1/2$ each of its n bits. This makes the probability of any given $\vec{r} = 1/2^n$, and the cost of generating the vector $\mathcal{O}(n)$.

Freivald's algorithm

The time complexity of Freivald's is $\Theta(n^2)$, we need time $\Theta(n^2)$ to compute the products $B \cdot \vec{r} =: \vec{r'}$, $A \cdot \vec{r'}$ and $C \cdot \vec{r}$, and additional time $\Theta(n)$ to check the equality (or not) of the vectors.

If $AB = C$ the algorithm yields always the correct answer. It could be that $AB \neq C$ and the algorithm yields a wrong answer ($AB = C$) with a certain probability. For example, with prob.=$1/2^n$, if we choose $\vec{r} = (0, 0, \ldots, 0)$ we will get the answer **true** even if $AB \neq C$.

Freivald's algorithm

Theorem

If $AB \neq C$ *then*

$$\mathbb{P}[A \cdot (B \cdot \vec{r}) = C \cdot \vec{r}] \leqslant \frac{1}{2}$$

Proof

Neat trick: As $AB \neq C$ taking $D = AB - C$, then $D \neq \mathbf{0}$.
$\Rightarrow \exists d_{ij} \in D$ s.t. $d_{ij} \neq 0$. W.l.o.g. assume $d_{11} \neq 0$.
If $\exists \vec{r}$ s.t. $A \cdot (B \cdot \vec{r}) = C \cdot \vec{r}$ then $D \cdot \vec{r} = 0$.
$D \cdot \vec{r} = 0 \Rightarrow \sum_{j=1}^{n} d_{1j} r_j = 0$, but as $d_{11} \neq 0$ then

$$r_1 = \frac{-\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}.$$

Freivald's algorithm

Theorem

If $AB \neq C$ *then*

$$\mathbb{P}[A(B(\vec{r})) = C\vec{r}] \leqslant \frac{1}{2}$$

Proof

Second trick: Choose $\vec{r} = (r_1, \ldots, r_n)$ from r_n to r_1 and stop at r_2, just before choosing r_1, which could be only 0 or 1.
Then the equality $r_1 = \frac{-\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$ holds with prob. $\leqslant 1/2$ □

Notice that by considering $r_n, \ldots, r_2$ to be fixed, we reduce the sample space to $r_1 \in \{0, 1\}$

Randomized algorithms and amplification

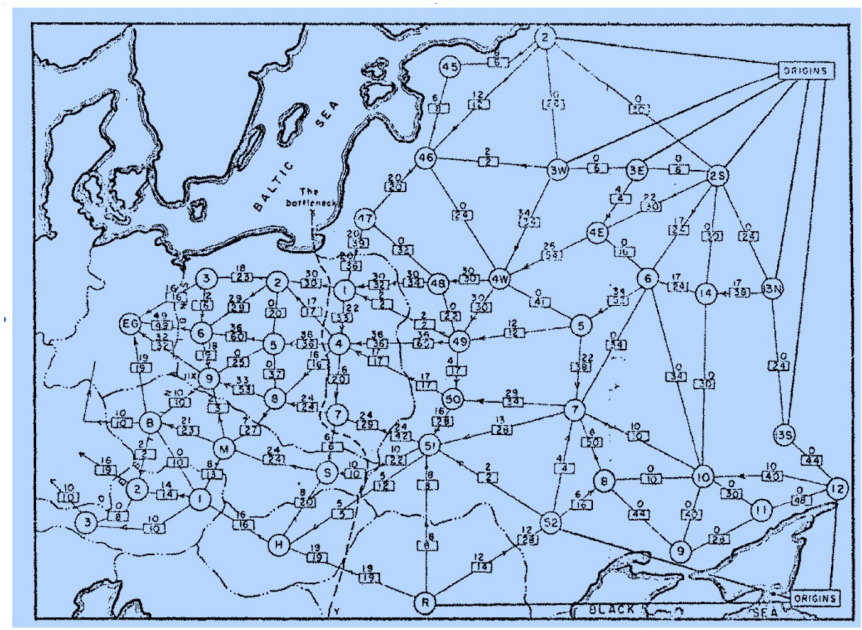
Freivald's algorithm finishes always in finite time ($\Theta(n^2)$) but may output the wrong answer. That type of randomized algorithms are called Monte Carlo algorithms.
Freivald's algorithm is also a one-side error algorithm, if $AB = C$ we always get the correct answer, but if $AB \neq C$ we may get the wrong answer with "small" probability.

One-side error Monte Carlo algorithms have the nice characteristic that can be amplified: each run of the algorithm can be considered as an independent "experiment", so they can be repeated, at each run we generate a new random choice, and by independence, each run decreases the probability of error.

If we repeat k times Freivald's algorithm and each time we generate a new $\vec{r}$, and the answer of the algorithm is $AB = C$ all the time, the probability of error (that is, that in fact $AB \neq C$) is $\leqslant 1/2^k$.

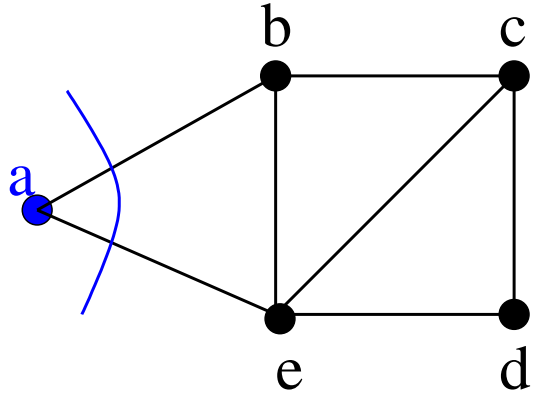
The Minimum Cut problem

In the mid 50's Harris and Ross studied the railway links between cities in the URSS and easter Europe and determined the easiest way to break the network by removing edges. The minimum cut of the graph.



The Minimum Cut problem

Given an undirected graph $G = (V, E)$ a cut is a partition of V in S and $\bar{S}$. The capacity of the cut is the number of edges with an end in S and the other in $\bar{S}$. The min cut is the cut with minimum capacity.



Complexity for deterministic algorithms

- Using Ford-Fulkerson: Max Flow-Min-Cut $\mathcal{O}(n^2 m)$ or $\mathcal{O}(nm)$ using J. Orlin's algorithms from 2013.
- Stoer-Wagner's algorithm (1994) $\mathcal{O}(nm + n^2 \lg n)$ (non-flow, weighted graphs)

Monte-Carlo algorithm for the Min-Cut problem

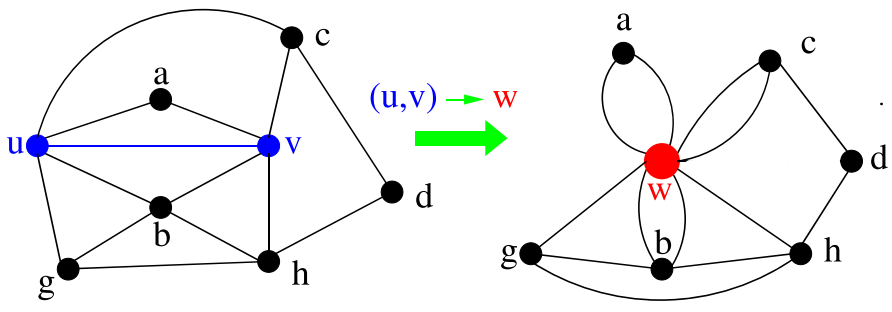
D. Karger, 1993.

Contracting an edge in G

Given a connected undirected graph $G = (V, E)$, we want to contract edges this operation will produce a graph with multiple edges but without self-loops:

procedure CONTRACT(G, $e = (u, v)$)
    Replace u and v by a super-node w
    Preserve edges, update endpoints of u and v to w
    Avoid self-loops but keep parallel edges
    return G
end procedure

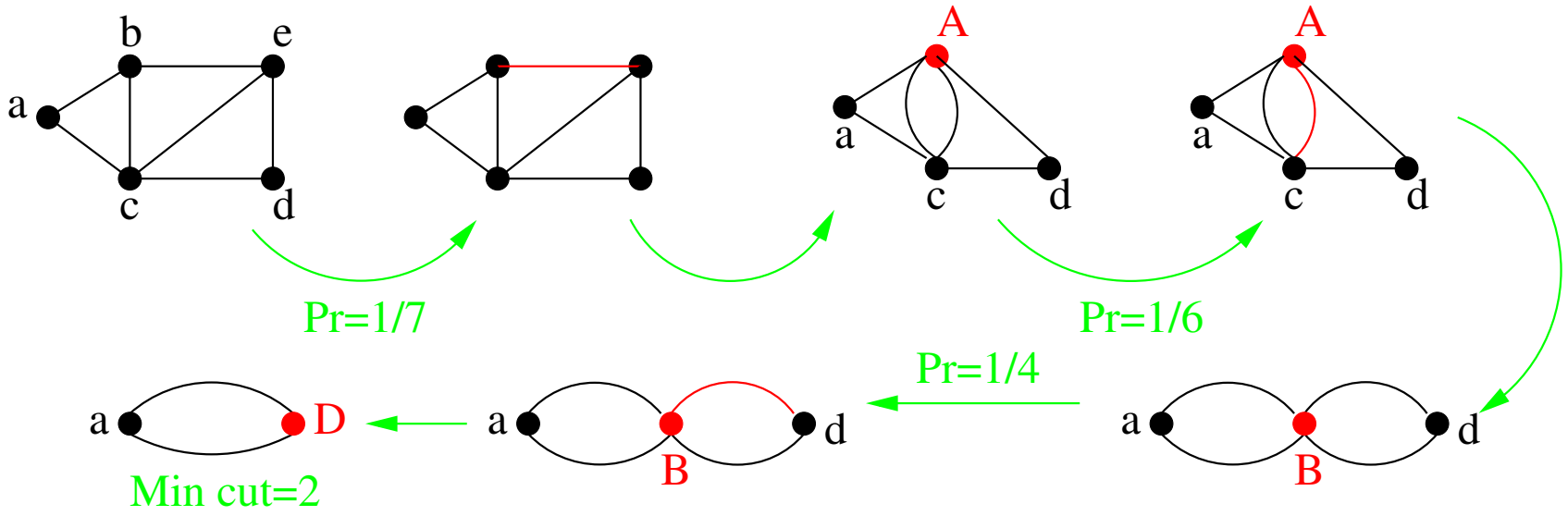
Monte-Carlo algorithm for the Min-Cut problem



Given G, which DS would you use to implement CONTRACT(e)?

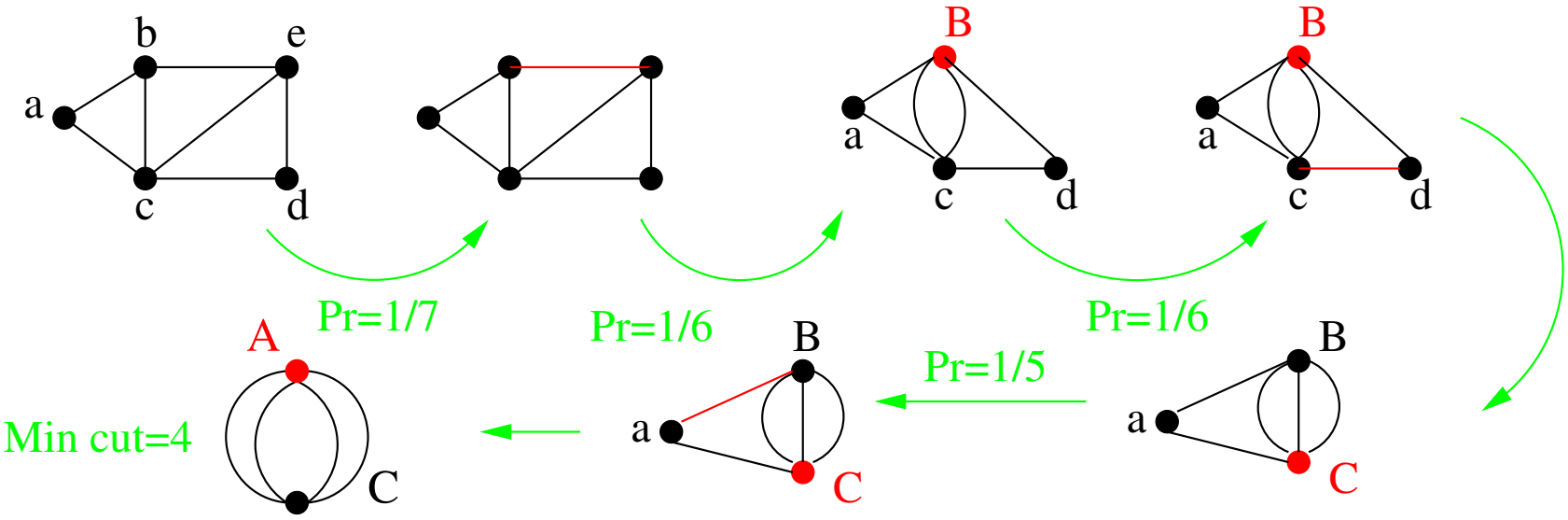
Karger's algorithm

procedure KARGER($G = \langle V, E \rangle$)
  while $|V| > 2$ **do**
    Choose u.a.r. $e = (u, v) \in E$
    $G := \text{CONTRACT}(G, e)$
  end while
  return # of edges between the 2 remaining vertices
end procedure



Karger's algorithm

procedure KARGER($G = \langle V, E \rangle$)
  while $|V| > 2$ **do**
    Choose u.a.r. $e = (u, v) \in E$
    $G :=$ CONTRACT(G, e)
  end while
  return # of edges between the 2 remaining vertices
end procedure



Analysis of the algorithm

The running time of the algorithm is $\Theta(n^2)$.

Assume G, with $|V| = n$ has a min-cut set $C \subseteq E$ of size k.

Notice:

- Any cut in a contracted graph is a cut in the initial graph,
- Karger's returns a cut,
- A contraction eliminates all the set of edges among the identified vertices.
- Karger's might provide a cut that is not of minimum size.

Theorem

Karger's algorithm returns a min-cut with probability $\geqslant 2/n^2$.

Proof of the Theorem

- Let C be a min-cut of G, assume that $|C| = k$
- Let G_i be the graph after i contractions, G_i has $n - i$ nodes.
- If no $e \in C$ has been contracted then C is still a min-cut of G_i, so $\Rightarrow |E(G_i)| \geqslant \frac{(n-i)k}{2}$ (as then, $\forall v \in V(G_i)$, v is adjacent to at least k edges)
- Let $\mathcal{E}_i$ be the event none of the edge(s) in C is contracted at the i-th iteration and let $\mathcal{F}_i = \cap_{j=1}^{i} \mathcal{E}_i$, i.e., no edge in C is contracted in the first i iterations.

Proof of the Theorem

- We want to compute $\mathbb{P}[\mathcal{F}_{n-2}]$, that is, probability of success.
- Notice $\mathbb{P}[\mathcal{E}_1] = \mathbb{P}[\mathcal{F}_1] \geqslant 1 - \frac{2k}{nk}$
  As $|C| = k$ all vertices in G must have degree $\geqslant k$, $|E(G)| \geqslant nk/2$. So 1st contracted edge chosen u.a.r. among the $\geqslant nk/2$ edges and $|C| = k$ choices to contract an edge in C
- $\mathbb{P}[\mathcal{E}_2|\mathcal{F}_1] \geqslant 1 - \frac{k}{k(n-1)/2} \geqslant 1 - 2/(n-1)$
  If 1st contraction did not eliminate an edge in C (i.e. conditioning on $\mathcal{F}_1$), we are left with $|V(G_1)| = n - 1$ and $|E(G_1)| \geqslant k(n-1)/2$, again because $\deg(v) \geqslant k$
- Working iteratively, $\mathbb{P}[\mathcal{E}_i|\mathcal{F}_{i-1}] \geqslant 1 - \frac{2}{(n-i+1)}$.

Proof of the Theorem

From $\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\,\mathbb{P}[B]$:

$$
\begin{aligned}
\mathbb{P}[\mathcal{F}_{n-2}] &= \mathbb{P}[\mathcal{E}_{n-2} \cap \mathcal{F}_{n-3}] \\
&= \mathbb{P}[\mathcal{E}_{n-2} \,|\, \mathcal{F}_{n-3}]\,\mathbb{P}[\mathcal{F}_{n-3}] \\
&= \mathbb{P}[\mathcal{E}_{n-2}|\mathcal{F}_{n-3}]\,\mathbb{P}[\mathcal{E}_{n-3}|\mathcal{F}_{n-4}] \cdots \mathbb{P}[\mathcal{E}_2|\mathcal{F}_1]\,\mathbb{P}[\mathcal{F}_1] \\
&\geqslant \prod_{i=1}^{n-2} \left(1 - \frac{2}{n-i+1}\right) = \prod_{i=1}^{n-2} \left(\frac{n-i-1}{n-i+1}\right) \\
&= \left(\frac{n-2}{n}\right)\left(\frac{n-3}{n-1}\right)\left(\frac{n-4}{n-2}\right) \cdots \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) \\
&= \frac{2}{n(n-1)}
\end{aligned}
$$

□

Amplification

To increase the probability of success, run Karger's algorithm several times.

Theorem

Run Karger's min-cut algorithm $n(n-1)\ln n$ *times and output the smallest cut found in all the runs.*
The probability of failure (it is not the global min-cut) is smaller or equal to

$$\left(1-\frac{2}{n(n-1)}\right)^{n(n-1)\ln n} \leqslant e^{-2\ln n} = \frac{1}{n^2}.$$

The proof is straightforward using the definition of e^{-1}