1 Introduction.

The theory of algorithms has undergone an extraordinary development in the last 20 years. The concept of "algorithm" is central in computer science. However, our aim is not to teach our students just a collection of algorithms. We want to develop the foundamental principles underlaying efficient algorithms and their analysis. Thus, each algorithm is developed starting from an abstract idea and further unfolded, with hopefully active participation from the reader, towards greater details and understanding just why each consecutive step is necessary or what other options are available.

Special attention will be given to understanding of the power that comes from applying probability in the teory of algorithms, to seeing the variety of ways in which probability plays a role. One usefull step in understanding this variety comes from making a clear distinction between the subject of probabilistic algorithms and probabilistic analysis of a (possibly deterministic) algorithm.

To begin with (Section 1), some tools of the trade in analysis of the algorithms will be discussed. This discussion will be followed (Section 2) by techniques for proving upper and lower bounds of algorithms and problems. When discussing the complexity of algorithms, special attention goes to average complexity (i.e. how does algorithm usually behave). Next (Section 3), few sorting techniques and sorting networks will be presented. in Section 4, we discuss some basics about random graphs. Sections 5-7 contain three algorithms based on different paradigms on which we will demonstrate our approach.

Our hope is that by asking questions, doing homework and commenting on it, correcting mistakes etc, our students will over time help us develop this subject further and produce the best learning tool in the area of algorithms.

2 Tools of the trade

We use algorithms to solve problems. A problem P (for example finding the maximum of a set of numbers) consists of infinite number of problem instances (one instance of a maximum problem is, for example, compute the maximum of 3,17,4,11,9). With every problem instance $p \in P$ one can associate a natural number representing the size of p (for our example maximum problem size was 5). The way the size is chosen for a given problem is not unique (we could have chosen to consider the total number of digits in our maximum problem, then the size would have been 7), but usually there is a natural choice.

Execution of the algorithm on a machine requires resources, most important ones being the run time of an algorithm and the amount of space (memory) it takes. But since each problem has many instances, it is generally a lot more informative to know about resource requirements of an algorithm, than to know what resources a particular instance requires (even though the later one may be also of interest in some cases). Global information such as run time for an algorithm on an input size n can not be determined by experiment. Two abstractions are generally used: worst case and average case behaviour. These are topics usually discussed in the course on the complexity of algorithms.

One can also talk about complexity of problems. An upper bound on the complexity of a problem is established by **designing and analysing** an algorithm that solves the problem. For example, a problem P is quadratic if there is an algorithm for P whose run time is bounded by a quadratic function. Nontrivial lower bounds are much harder to establish. To show that a lower bound is quadratic for some problem P, one must show that every algorithm that solves P has at least quadratic run time.

Our goal is to fully understand some paradigms based on which algorithms can be designed and to have methods for analysing and comparing so produced algorithms.

It is very common that the analysis of algorithm produces a sum, usually finite, for which no representation in primitive terms or known functions exists. Or, one can come up with representation involving primitive terms or known functions, but the answer is too complicated for all practical purposes. The most common method for dealing with such situations is asymptotic approximation.

2.1 Asymptotic Notation

Asymptotic analysis is concerned with providing approximations that are useful when some parameter becomes large. The O, Θ and Ω notations lead to results that are something in between saying that one function is approximately equal to another and setting an upper and a lower bound on a function. Using this notation requires less work than computation of exact bounds, but it does not give the information on how big is the error at any point.

The following are the definitions for big O, big Θ, Ω and little o:

- $O(f(x))$ is the set of all functions $g(x)$ such that there exist positive constants C and x_0 with $g(x) \leq Cf(x)$ for all $x \geq x_0$, i.e. $f(x)$ provides an upper bound for the set of functions $g(x)$, up to a multiplicative constant.
- $\Theta(f(x)$ is the set of all functions $g(x)$ such that there exist positive constants C, C' and x_0 with $Cf(x) \leq g(x) \leq C'f(x)$ for all $x \geq x_0$, i.e. $f(x)$ provides a lower bound, up to a multiplicative constant.
- Ω is the set of all functions $g(x)$ such that there exist positive constants C and x_0 with $g(x) \geq Cf(x)$ for all $x \geq x_0$, i.e. function $f(x)$ and $g(x)$ behave roughly the same way for large x.
- o is the set of all functions $g(x)$ such that $\lim_{x \Rightarrow \infty} \frac{g(x)}{f(x)} = 0$, i.e. for large x $g(x)$ is much smaller than $f(x)$.

The function $f(x)$ is assumed to be nonnegative.

These notation simplify relations, allowing one to concentrate on dominant terms. But one should never forget that equality signs used with this notations are not the actual equalities. Thus term error notation is also used for asymptotic notation.
Here are some examples of the usage:

$$2n^2 = O(n^2)$$
$$2n^2 = \Omega(n)$$

$$2n^2 + n = \Theta(n^2)$$
$$2n^2 = o(n^2 lnn)$$

More details about asymptotic notation can be found in [**?**] and [**?**].

2.2 Asymptotic Approximation

Before launching into specific tools and examples, we will try to set the stage a bit and give an intuitive and descriptive approach to most common situations in which asymptotic estimates are required.

Suppose we are interested in a sequences of numbers. We have four basic methods for providing information about those numbers.

- **A combinatorial description**: say P(n) is the number of partitions in an n-set;
- **A formula**: the number of involutions of an n set is

$$\sum_{j=0}^{n} \frac{n!}{j!2^j(n-2j)!};$$

- **A recursion**: $F_0 = 1$, $F_1 = 2$ and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$;
- **A generating function**: the ordinary generating function for the number of comparisons needed to Quicksort an n long list is

$$\frac{-2\ln(1-x) - 2x}{(1-x)^2}.$$

Given such information, can we obtain some information about the size of the terms in the sequence? The answer will depend on the information we are given. Generally, the situation will be as follows:

- **A combinatorial description**: It is usually difficult, if not impossible to obtain information about the size of the terms from such a description.

- **A formula**: The formula by itself may be explicit enough. If it is not, using Stirling's approximation formula may help. Calculating upper and lower bound may give a very good answer. If summation is involved, most likely it can be estimated using some of the techniques disscused below.

- **A recursion**: It is often possible to obtain some information. A simple case will be discussed later.

- **A generating function**: If the generating function converges for some values of x different than $x = 0$, it is likely that some tools from analysis can be used to estimate the coefficients. In this section we will discuss only some very simple tools.

However, there is generally no simple answer to the the problem of finding the approximation to some $f(n)$ for large values of n. We must ask how simple and how accurate approximation is desired.

We will not specify precisely what constitutes a simple expression, but the following example may give you the feel for it. The expression $\sqrt{2\pi n}(\frac{n}{e})^n$ is simpler than the expression $n!$ even though the later is easier to write. Why? If we limit ourselves to basic operations (addition, multiplication, division and exponentiation) then the former expression requires the constant number of operations (six), while the later requires $n - 1$ operations.

There are wide variations in the degree of accuracy that we might ask for. Generally, we would like an approximation such that the relative error goes to zero. In other words, given some $f(n)$ we want to find $Apx(f_n)$ such that $\frac{|f(n)-Apx(f_n)|}{Apx(f_n)} \to 0$ as $n \to \infty$. Unfortunately, even when we know that relative error goes to zero eventually, we do not know what eventually means. It can take a very long time. Finding an upper and a lower bound is the answer even though it can sometimes be too difficult to compute and even if one has a bound it can be overly pessimistic.

Short formulas for upper and lower bounds on a function will, if close together, contain almost the same information about the size of the function as an exact formula for the value of the function.

Most upper and lower bound calculations can be done using four groups of principles. The first group is:

1. In an upper bound one can replace any quantity by another quantity which is known not to be smaller.

2. In a lower bound one can replace any quantity by another quantity which is known not to be larger.

For **example**, some Linear Search Algorithm has average number of steps $\frac{1}{2}n + \frac{1}{2}$. One can obtain a lower bound by dropping the term $\frac{1}{2}$. One can obtain the upper bound for $n \geq 1$ by replacing $\frac{1}{2}$ with $\frac{1}{2}n$. Thus,

$$\frac{1}{2}n \leq \frac{1}{2}n + \frac{1}{2} \leq n.$$

The second group of principles tells us how to combine upper and lower bounds arithmetically. Let $f(x)$ and $g(x)$ be two functions with respective upper and lower bounds $f_U(x), f_L(x), g_U(x), g_L(x)$ on some closed interval.
3. The sum of upper (lower) bounds is an upper (lower) bound on the sum:

$$f_L(x) + g_L(x) \leq f(x) + g(x) \leq f_U(x) + g_U(x)$$

4. $cf_L(x) \leq cf(x) \leq cf_U(x)$ when $c \geq 0$ **5.** $cf_U(x) \leq cf(x) \leq cf_L(x)$ when $c \leq 0$ **6.** Multiplying a nonnegative upper (lower) bound by a nonnegative upper (lower) bound gives an upper (lower) bound on the product of functions:

$$f_L(x)g_L(x) \leq f(x)g(x) \leq f_U(x)g_U(x)$$

when all the functions are positive.

The next group of principles tells how upper and lower bounds behave with increasing and decreasing functions. Frequently encountered representatives of increasing functions in the analysis of algorithms are exponential and logarithmic functions and of decreasing functions the reciprocal, $\frac{1}{x}$ $(x \neq 0)$.
7. Let $f(y)$ be an increasing function in the range

$$\min_{x_0 \leq x \leq x_1} g_L(x) \leq \max_{x_0 \leq x \leq x_1} g_U(x).$$

Then, for $x_0 \leq x \leq x_1$

$$f(g_L(x)) \leq f(g(x)) \leq f(g_U(x)).$$

8. Let $f(y)$ be a decreasing function in the range

$$\max_{x_0 \leq x \leq x_1} g_U(x) \leq \min_{x_0 \leq x \leq x_1} g_L(x).$$

Then, for $x_0 \leq x \leq x_1$

$$f(g_U(x)) \leq f(g(x)) \leq f(g_L(x)).$$

For **example**, in a specific random hashing algorithm, the average number of times a certain step is performed is $1 + \frac{k}{(N-k)}$, where $N > k$. If we consider k in the range $0 \leq k \leq \frac{N}{2}$, we have $\frac{N}{2} \leq N - k \leq N$ by using Principles 3 and 5. B y principle 8 we obtain $\frac{k}{N} \leq \frac{k}{(N-k)} \leq \frac{2k}{N}$. Finally, using Principle 1, we obtain

$$1 + \frac{k}{N} \leq 1 + \frac{k}{(N-k)} \leq 1 + \frac{2k}{N}.$$

These eight principles work for simple examples. In more complex cases, one often needs an additional principle based on Taylor's theorem.

Theorem 2.1 *If $f(x)$ is a function with a continuous n^{th} derivative on the closed interval [a,b], and x and x_0 are two distinct points of [a,b], then there exists a point c between x and x_0 such that*

$$f(x) = \sum_{0 \leq i \leq n} \frac{f^i(x_0)}{i!}(x - x_0)^i + \frac{f^{n+1}(c)}{(n+1)!}(x - x_0)^{n+1}.$$

Proof: Let $P_n(x)$ be Taylor's polynomial (i.e. polynomial approximation to $f(x)$) and let M be a number defined by

$$f(\alpha) = P_n(\alpha) + M(\alpha - x_0)^{n+1}, \alpha \in [a, b], \alpha \neq x_0$$

and let

$$g(x) = f(x) - P_n(x) + M(x - x_0)^{n+1}, \quad a \leq x \leq b.$$

We have to show then that $(n+1)!M = f^{n+1}(c)$ for some c between x_0 and α. Taking $n+1$ derivatives of $g(x)$ we obtain:

$$g^{n+1}(x) = f^{n+1}(x) - (n+1)!M, \quad a < x < b.$$

The proof is then complete if we show $g^{n+1}(x) = 0$ for some x between x_0 and α. Since $P^k(x_0) = f^k(x_0) \ \ \forall k = 1, \ldots, n$, we have

$$g(x_0) = \ldots = g^n(x_0) = 0.$$

Our choice of M shows that $g(\alpha) = 0$, so that $g'(x_1) = 0$ for some x_1 between x_0 and α. (This follows from the mean value theorem that postulates that if $g(x_0) = g(\alpha) = 0$, then $g'(x_1) = 0$ for some x_1 between x_0 and α. Since $g'(\alpha) = 0 \Rightarrow g''(x_2) = 0$ for some x_2 between x_1 and α (by applying the mean value theorem again). After n steps $g^{n+1}(x_{n+1}) = 0$ for some x_{n+1} between x_0 and x_n, i.e. between x_0 and α.

Example. Use the third degree Taylor's polynomial to approximate the function $\ln x$. Calculate the value of the polynomial at point 1.1 (i.e. approximate the value of $\ln 1.1$) and estimate the error.

We first find Taylor's polynomial of degree three centred at $x = 1$.

$$f(x) = \ln x \qquad f(1) = \ln 1 = 0$$

$$f'(x) = \frac{1}{x} \qquad f'(1) = 1$$

$$f'''(x) = \frac{2}{x^3} \qquad f''(1) = 2$$

$$P_n = \sum_{0 \le i \le n} \frac{f^i(x_0)}{i!}(x - x_0)^i = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3$$

Now the polynomial is simply evaluated at point 1.1 obtaining the result $P_3(1.1) \approx 0.0953$. The error term is given by $\frac{f^{n+1}(c)}{(n+1)!}(x - x_0)^{n+1} = \frac{f^4(c)}{4!}(x - 1)^4 = -\frac{1}{4c^4}(x-1)^4$. Let $x = 1.1$. We obtain $R_3(1.1) = -\frac{1}{4c^4}10^{-4}$. The maximum of this function on the interval $1 \le c \le 1.1$ is $\frac{1}{4}$ at $c = 1$ and thus $|R_3(1.1)| \le 2.510^{-5}$.

Problem a) Give a third degree polynomial, centred at 1, that approximates the function $x^{\frac{1}{3}}$. b) Use this polynomial to give approximate value of $(1.4)^{\frac{1}{3}}$. c) Estimate the error in your approximation.

The principle for bounding the function using Taylor's theorem can thus be stated as:

9.

$$f(x) \ge \sum_{0 \le i \le n} \frac{f^i(x_0)}{i!}(x - x_0)^i + L(x - x_0)^{n+1}.$$

$$f(x) \leq \sum_{0 \leq i \leq n} \frac{f^i(x_0)}{i!}(x - x_0)^i + U(x - x_0)^{n+1}.$$

where U and L respectively are upper and lower bound on $\frac{f^{n+1}(x)}{(n+1)!}$ in the range of x.

2.3 Euler's Summation Formula

The Euler's summation formula gives a difference between an integral and the corresponding sum. To derive the formula, we consider the area under the curve, on the desired interval. Interval is then split into n subintervals and we compare the area of one trapezoid (with the base on subinterval $[i, i+1]$ and the remaining two vertices $(i, f(i))$ and $(i+1, f(i+1))$) and the area under the corresponding part of the curve. If we integrate $(x - i - \frac{1}{2})f'(x)$, we obtain

$$\int_i^{i+1} (x - i - \frac{1}{2})f'(x)dx = (x - i - \frac{1}{2})f(x) \Big|_{i}^{i+1} - \int_i^{i+1} f(x)dx$$

$$= \frac{1}{2}[f(i+1) + f(i)] - \int_i^{i+1} f(x)dx$$

That is, our integral is the difference between the area of trapezoid and the exact area under the curve on subinterval $[i, i+1]$.

We would like now to relate the integral for the entire area to the value of the function at integer points. First, we rewrite $x - i - \frac{1}{2}$ in a way that removes the i dependence. We define the *sawtooth function* $\{x\}$ as

$$\{x\} = x mod 1 = x - \lfloor x \rfloor.$$

The integral can now be rewriten as

$$\int_i^{i+1} (\{x\} - \frac{1}{2})f'(x)dx = \frac{1}{2}[f(i+1) + f(i)] - \int_i^{i+1} f(x)dx.$$

Let us now sum over all subintervals, i.e. compute

$$\int_1^n (\{x\} - \frac{1}{2})f'(x)dx = \frac{1}{2}\sum_{i=1}^{n-1}[f(i+1) + f(i)] - \int_1^n f(x)dx$$

$$= \sum_{i=1}^{n-1} f(i) + \frac{1}{2}[f(n) - f(1)] - \int_1^n f(x)dx.$$

Last equation can now be used to compute integrals from sums or sums from integrals, provided one can estimate the size of the integral. Since we will generally be interested in computing sums from integrals, we can rewrite the above as

$$\sum_{i=1}^{n-1} f(i) = \int_1^n f(x)dx - \frac{1}{2}[f(n) - f(1)] + \int_1^n B_1(\{x\})f'(x)dx$$

where $B_1(x)$ is the polynomial $x - \frac{1}{2}$.

Example. Approximate $\sum_{i=1}^{n-1} \frac{1}{i}$.

$$\sum_{i=1}^{n-1} \frac{1}{i} = \int_1^n \frac{1}{x}dx - \frac{1}{2}(\frac{1}{n} - 1) - \int_1^n B_1(\{x\})\frac{1}{x^2}dx$$

$$= \ln x + \frac{1}{2} - \frac{1}{2}n - \int_1^n B_1(\{x\})\frac{1}{x^2}dx$$

An approximation to $\int_1^n B_1(\{x\})\frac{1}{x^2}dx$ is now needed in order to obtain the approximation for the sum. We observe that $B_1(\{x\})$ is between $-\frac{1}{2}$ and $\frac{1}{2}$ for any x, so we have

$$-\frac{1}{2}\int_1^n \frac{1}{x^2}dx \leq \int_1^n B_1(\{x\})\frac{1}{x^2}dx \leq \frac{1}{2}\int_1^n \frac{1}{x^2}dx$$

or

$$-\frac{1}{2}(1 - \frac{1}{n} \leq \int_1^n B_1(\{x\})\frac{1}{x^2}dx \leq \frac{1}{2}(1 - \frac{1}{n}dx.$$

Therefore, $\sum_{i=1}^{n-1} \frac{1}{i} = \ln n + O(1)$. However, a better approximation can be obtained by integrating $\int_1^n B_1(\{x\})\frac{1}{x^2}dx$ by parts again. As this process of integration by parts is continued, a sequence of *Bernoulli polynomials* is obtained. The Bernoulli polynomial $B_m(x)$ is defined recursively as $\int mB_{m-1}(x)dx$ with integration constant equal to 1. The *Bernoulli numbers* B_m, which are coefficients of Bernoulli polynomials, are defined by the recurrence

$$B_0 = 1, \quad B_n = \frac{-1}{n+1}\sum_{i=1}^{n-1} \binom{n+1}{i} B_i, \quad n \geq 1.$$

Except for B_1 all the Bernoulli numbers of odd index are zero. The first few Bernoulli numbers are: $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_3 = 0$, $B_4 = -\frac{1}{30}$, $B_5 = 0$, $B_6 = \frac{1}{42}$, $B_8 = -\frac{1}{30}$, $B_{10} = \frac{5}{66}\ldots$.

The Bernoulli polynomials are then given by

$$B_m(x) = \sum_{i=0}^{m} \left(\begin{array}{c} m \\ i \end{array} \right) B_i x^{m-i}.$$

For $m \geq 1$, integration by parts yields

$$\int_1^n B_m(\{x\}) f^m(x) dx = \frac{1}{m+1}(B_{m+1} f^m(n) - B_{m+1} f^m(1) - \int_1^n B_{m+1}(\{x\}) f^{m+1}(x) dx).$$

Repeated application of the last equation gives so called *Euler's general summation formula:*

$$\sum_{i=1}^{n-1} f(i) = \int_1^n f(x)dx + \sum_{i=1}^{m} \frac{B_i}{i!}(f^{i-1}(n) - f^{i-1}(1)) + R_m$$

where

$$R_m = \frac{(-1)^{m+1}}{m!} \int_1^n B_m(\{x\}) f^m(x) dx).$$

The remainder R_m will be small when $\frac{B_m(\{x\}) f^m(x)}{m!}$ is small. It is known that $|B_m(\{x\})| \leq |B_m|$ when m is even and that

$$\left| \frac{B_m(\{x\}) f^m(x)}{m!} \right| \leq \left| \frac{4}{(2\pi)^m} \right|$$

so $\frac{B_m(\{x\}) f^m(x)}{m!}$ does not cause any trouble. On the other hand for many functions $f^m(x)$ becomes large for large m. In such cases, for fixed x, there is a best value of m at which R_m has its minimum. This will not be dealt with here.

When Euler's formula is applied to slowly varying functions, the error term rapidly decreases as the order of approximation increases. However, if function values vary rapidly for small values, one must be careful in applying the formula directly. An example or rather rapidly varying function for small values of the argument is function $f(i) = \frac{1}{i}$. The sum $\sum_{i=1}^{n} \frac{1}{i}$ was mentioned earlier. It is called the *n-th Harmonic number*

$$H_n = \sum_{i=1}^{n} \frac{1}{i}.$$

We have applied Euler's formula to this function and have obtained the result $H_n = \ln n + O(1)$. Let's see what result gives the application of Euler's general summation formula. Substituting $f(x) = \frac{1}{x}$ and $f^m(x) = \frac{(-1)^m m!}{x^{m+1}}$ into general formula yields:

$$H_{n-1} = \ln n + \sum_{i=1}^{m} \frac{B_i}{i}(-1)^{i-1}(\frac{1}{n^{(i)}} - 1) + R_m.$$

The error term here however, is not decreased as m increases. One can remedy this by considering the value of the error as n becomes large and subtract off the limiting value as follows:

$$\gamma = \lim_{n\to\infty}(H_{n-1} - \ln n) = \sum_{i=1}^{m} \frac{B_i}{i}(-1)^i - \int_1^\infty \frac{B_m(\{x\})}{x^{m+1}}.$$

The constant γ is called *Euler's constant* and its value is 0.57721... Now,

$$\sum_{i=1}^{m} \frac{B_i}{i}(-1)^i - \sum_{i=1}^{m} \frac{B_i}{i}(-1)^{i-1}(\frac{1}{n^i} - 1) = \sum_{i=1}^{m} \frac{B_i}{in^i}(-1)^i$$

and

$$\int_1^\infty \frac{B_m(\{x\})}{x^{m+1}}dx - R_m = \int_n^\infty \frac{B_m(\{x\})}{x^{m+1}}dx.$$

Thus,

$$H_{n-1} = \ln n + \gamma + \sum_{i=1}^{m} \frac{B_i}{in^i}(-1)^i + \int_n^\infty \frac{B_m(\{x\})}{x^{m+1}}dx.$$

Evaluating the last formula for $m = 8$ and adding $\frac{1}{n}$ to both sides gives

$$H_n = \ln n + \gamma + \frac{1}{2n} + \frac{1}{12n^2} + \frac{1}{120n^4} + \frac{1}{256n^6} + O(\frac{1}{n^8}).$$

The most common way that harmonic numbers arise in algorithm analysis is from finding the largest of n numbers. When one looks at the numbers from the first to last, the probability that the first is the largest is 1, the probability that the second is largest is 0.5, etc. The sum of these probabilities gives the expected number of new maximums, so the average number of new maximums is H_n.

2.4 Stirling's Approximation

To approximate the factorial function $n!$, one can start by looking at the logarithm of $n!$. Since $n! = n(n-1)(n-2)\dots 21$, $\ln n! = \ln n + \ln(n-1) + \ln(n-2) + \dots + \ln 2 + \ln 1$. Since $n!$ is a rapidly varying function, Euler's summation formula can not be applied to it directly. But $\ln n!$ is a slowly varying function, so Euler's summation formula can be used on it. An accurate approximation of the logarithm of $n!$ leads to an approximation for $n!$ itself which has small relative error, but not necessarily a small absolute error. Using $f(x) = \ln x$, we have $f^m(x) = \frac{(-1)^{m+1}(m-1)!}{x^m}$ and $\int_1^x f(y)dy = x\ln x - x + 1$. Eulers general summation formula then gives

$$\ln(n-1)! = n\ln n - n + 1 - \frac{1}{2}\ln n + \sum_{1<i\leq m} \frac{B_i(-1)^i}{i(i-1)}(\frac{1}{n^{i-1}} - 1) + R_m,$$

where

$$R_m = \frac{1}{m}\int_1^n \frac{B_m(\{x\})}{x^m}dx \quad for\ \ m \geq 2.$$

Again, as we did with harmonic numbers, we look at the limit as n approaches infinity of

$$\lim_{n\to\infty}(\ln n! - n\ln n + n - \frac{1}{2}\ln n) = 1 + \sum_{1<i\leq m} \frac{B_i(-1)^{i+1}}{i(i-1)} + \lim_{n\to\infty} R_m.$$

One should notice here that R_m really is a function of n as well and that this limit exists. Therefore, let us give a name to this limit, i.e. let $\lim_{n\to\infty} R_m = \sigma$. Without computing the values of the limits in a moment, we can now write

$$\ln n! = (n + \frac{1}{2})\ln n - n + \sigma + \sum_{1<i\leq m} \frac{B_i(-1)^i}{i(i-1)} + O(\frac{1}{n^m}).$$

Setting $m = 3$ and taking exponentials gives

$$n! = e^\sigma \sqrt{n}(\frac{n}{e})^n e^{(\frac{1}{12n} + O(\frac{1}{n^3}))}.$$

We should now find the value of σ. Let us consider the expression $\frac{\sqrt{n}(2n)!}{4^n n!n!}$. Evaluating this expression using the above formula for $n!$, we get

$$\frac{\sqrt{n}(2n)!}{4^n n!n!} = \frac{\sqrt{2}}{e^\sigma}(1 + O(\frac{1}{n}))$$

so that the limit as n approaches infinity of this expression is $\frac{\sqrt{2}}{e^\sigma}$. If another way of evaluating the limit can be found, then σ can be computed. So we try

$$\frac{\sqrt{n}(2n)!}{4^n n!n!} = \sqrt{n}\frac{1 \cdot 2 \ldots 2n}{2 \cdot 4 \ldots 2n \cdot 2 \cdot 4 \ldots 2n}.$$

If we square this expression, we obtain

$$n(\frac{(2n)!}{4^n n!n!})^2 = \frac{n}{2n+1} \prod_{1 \leq i \leq n} (\frac{(i-\frac{1}{2})(i+\frac{1}{2})}{i \cdot i}).$$

If we take the limit as n goes to infinity of this last expression, we obtain Gamma functions on the right hand side, and knowing their values, we obtain:

$$\lim_{n \to \infty} n(\frac{(2n)!}{4^n n!n!})^2 = \frac{\Gamma(1)\Gamma(1)}{2\Gamma(\frac{1}{2})\Gamma(\frac{3}{2})} = \frac{1}{\pi}.$$

From this we can now easily find that

$$e^\sigma = \sqrt{2\pi}$$

and so, finally, obtain Stirling's approximation formula

$$n! = \sqrt{2\pi n}(\frac{n}{e})^n e^{(\frac{1}{12n} + O(\frac{1}{n^3}))}.$$

This approximation will be good enough for our purposes. However, we remark that better approximations can be obtained by expanding the exponential in a power series.

2.5 Approximating Binomial Coefficients and some Basic Counting Techniques

In combinatorics, binomial coefficients in importance come right after natural numbers. They appear in various contexts and diguises. Usually, the first encounter with them is through so called binomial theorem

$$(a+b)^n = \sum_{i=0}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) a^i b^{n-i}.$$

Combinatorial interpretation of binomial coefficients is the famous counting the number of ways in which one can select k distinct objects out of the set of n objects. This can also be viewed as the number of k-permutations taken from an n set in which order does not matter, so we have

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{P(n,k)}{P(k,k)} = \frac{n!}{(n-k)!k!}$$

Using Stirling's formula, one can obtain the following asymptotic values for binomial coefficients

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \left(\frac{n}{2\pi k(n-k)} \right)^{\frac{1}{2}} \left(\begin{array}{c} n \\ k \end{array} \right)^{k} \left(\begin{array}{c} n \\ n-k \end{array} \right)^{n-k} (1+\frac{1}{12}(\frac{1}{n}-\frac{1}{k}-\frac{1}{n-k})+O(\frac{1}{kn(n-k)})).$$

If $j = \frac{n}{2} - k = o(n)$, then

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \left(\frac{2}{\pi n} \right)^{\frac{1}{2}} 2^n e^{\frac{-2j^2}{n}}.$$

Thinking in the counting context, we can relax the condition that objects are distinct, and consider having n objects, n_1 of which are of one kind (say red), n_2 of the second kind (blue) etc. Then the number of permutations (with duplication) is so called multinomial coefficient

$$\left(\begin{array}{c} n \\ n_1 n_2 \dots n_k \end{array} \right) = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Example. How many walks are there in a sqare lattice that begin and end in origin?
It is easy to see that the number of steps up must be the same as the number of steps down, as well as that the number of steps to the left must be the same as the number of steps to the right and that total number of steps has to be even, let's say $2n$. Let $W(k)$ be the number of walks in which we make k steps up. Then we also make k steps up, and we make $n-k$ steps to the left, as well as to the right. Then

$$W(k) = \frac{(2n)!}{k!^2 (n-k)!^2}$$

The final answer is the sum over all possible numbers of steps up $\sum_{k=0}^{n} W(k)$.

In a similar vein, one can consider combinations (selections) with repetitions allowed. Imagine that we have n objects and want to select k of them, with freedom to select the same object more than once. Or, equivalently, as having objects of n different types, each type in unlimmited supply (acctually k copies of each suffice). In how many ways can we select k of these objects? The answer to this question can be found quite elegantly by establishing bijection between our formulation and finding integer solutions to equation $x_1 + x_2 + \ldots + x_n = k$. In fact, we have

Theorem 2.2 *Let k and n be naturl numbers. Then the number of ways to choose k objects from a collection of n distinct objects with repetition allowed is equal to the number of solutions in natural numbers to the equation $x_1 + x_2 + \ldots + x_n = k$ and that number is*

$$\left(\begin{array}{c} n+k-1 \\ k \end{array} \right).$$

Proof: we will establish a one-to-one correspondence between combinations and solutions to the equation. To specify a selection of k objects from the set of n objects with repetition allowed, it sufice to specify how many times the first object is picked, how many times the secon and so on. Thus the selection is uniquely determined and determines a n-tuple of integers $(x_1, x_2, \ldots, x_n)$ such that each $x_i \geq 0$ and $x_1 + x_2 + \ldots + x_n = k$. It remains to determine the number of solutions to the equation. Now, consider all strings of length $n + k - 1$ consisting of k $*$ and $n - 1$ $\backslash$. For example

$$\backslash * \backslash * * \backslash\backslash * * * \backslash\backslash\backslash \qquad (k = 6, n = 7).$$

It is easy to see that the cardinality of this set is $\left(\begin{array}{c} n+k-1 \\ k \end{array} \right)$ (there are $n + k - 1$ positions, k of them to be filled with $*$). Correspondence between strings and solutions to the equation is a simple one: the number of $*$ before the first $\backslash$ corresponds to x_1, between first and second $\backslash$ to x_2 and so on. This completes the proof.

Example. Determine the number of ordered partitions of positive integer k into n positive parts. Let $x_1, x_2, \ldots, x_n$ be parts into which we want to partition. Each $x_i \geq 1$. let us then substitute x_i by $x_i' = x_i - 1$. Then we ask for the number of solutions to $x_1' + x_2' + \ldots + x_n' = k - n, \quad x_i' \geq 0$. It is now easy to see that the answer is

$$\left(\begin{array}{c} n + (k-n) - 1 \\ k - n \end{array} \right) = \left(\begin{array}{c} k-1 \\ k-n \end{array} \right) = \left(\begin{array}{c} k-1 \\ n-1 \end{array} \right).$$

Often one discovers interesting relations involving binomial coefficients by counting in two different ways.
Example. How many different committees with 4 or 5 members is it possible to appoint from US senate which has 435 members?
The streight forward answer is

$$\left(\begin{array}{c} 435 \\ 4 \end{array} \right) + \left(\begin{array}{c} 435 \\ 5 \end{array} \right).$$

Looking at the problem a bit differently, we can think of senate as having 436 members, one of which is fictive (a dummy). Now, all we want is 5 members committees from these 436 members. In whichever committee a dummy appears, we will regard it as a four member committee. Therefor, the answer is

$$\left(\begin{array}{c} 436 \\ 5 \end{array} \right).$$

This produces the identity

$$\left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n \\ k+1 \end{array} \right) = \left(\begin{array}{c} n+1 \\ k+1 \end{array} \right).$$

This identity is known as **Pascal's identity**. All sorts of interesting patterns, theorems and conjectures emerge from this identity. It is also very handy for demonstrating the logic of combinatorial, algebraic and inductive proving methods. So we will take a bit of digression and prove the identity using those three methods.
1. Pascal's identity, combinatorial proof.
To prove that the left hand side equals the right hand side, we will show that they count the same thing. The right hand side tells us how many subsets of

cardinality $k+1$ a set with $n+1$ elements has (basic definition of binomial coefficient). We will now show that the left hand side also counts subsets of size $k+1$ of an $n+1$ element set. Let us select a specific element from an $n+1$set. Let's call it x. Either a subset of $k+1$ elements will contain x, or it will not. These possibilities are disjoint. If we want to select a subset that does not contain x, then there are $\left(\begin{array}{c} n \\ k+1 \end{array}\right)$ ways to do this (i.e., without x set has n elements, and we want to select $k+1$ of them). If we do want to include x into our subset, then we need to select only k elements in addition to x, again from the remaining n elements. By addition principle, there are $\left(\begin{array}{c} n \\ k \end{array}\right) + \left(\begin{array}{c} n \\ k+1 \end{array}\right)$ ways to choose $k+1$ subsets of $n+1$ elements, and the proof is complete.

2. Pascal's identity, algebraic proof.

$$\left(\begin{array}{c} n \\ k \end{array}\right) + \left(\begin{array}{c} n \\ k+1 \end{array}\right) = \left(\begin{array}{c} n+1 \\ k+1 \end{array}\right).$$

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

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

We can now cancel appropriate terms on the left and right hand side, to obtain

$$\frac{1}{(n-k)} + \frac{1}{(k+1)} = \frac{(n+1)}{(n-k)(k+1)}$$

Now left and right hand sides are obviously identical (putting the left hand side to a common denominator gives the right hand side), and the proof is complete.

3. Pascal's identity, inductive proof.

The induction is based on k. Let's start by proving the base case $k=1$.

$$\left(\begin{array}{c} n \\ 1 \end{array}\right) + \left(\begin{array}{c} n \\ 2 \end{array}\right) = \left(\begin{array}{c} n+1 \\ 2 \end{array}\right),$$

$$n + \frac{n(n-1)}{2} = \frac{(n+1)n}{2}.$$

By finding the common denominator for the left hand side, we see that left and right hand sides are identical and therefor the base case is true. Let us suppose that (inductive hypothesis)

$$\left(\begin{array}{c} n \\ k-1 \end{array} \right) + \left(\begin{array}{c} n \\ k \end{array} \right) = \left(\begin{array}{c} n+1 \\ k \end{array} \right)$$

holds. From the hypothesis, we find that

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \left(\begin{array}{c} n+1 \\ k \end{array} \right) - \left(\begin{array}{c} n \\ k-1 \end{array} \right) \quad (*).$$

Now we want to prove (inductive step)

$$\left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n \\ k+1 \end{array} \right) = \left(\begin{array}{c} n+1 \\ k+1 \end{array} \right).$$

Let us use $(*)$ on $\left(\begin{array}{c} n \\ k+1 \end{array} \right)$ and substitute the result into the last equation. We get

$$\left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n \\ k+1 \end{array} \right) = \left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n+1 \\ k+1 \end{array} \right) - \left(\begin{array}{c} n \\ k \end{array} \right).$$

Cancellation of appropriate terms leads to the desired result.

2.6 Linear First Order Recurrences

The reader is expected to know already how to solve the simplest recurence relations. Our interest now is a bit different - we want asymptotic information from recursion. We will limit the discussion at this time to linear recurrences with "almost constant" coeficients. In other words, we will look at, except for initial conditions, at

$$a_n = c_1(n)a_{n-1} + c_2(n)a_{n-2} + \cdots + c_k(n)a_{n-k} \qquad (*)$$

where the functions $c_i(n)$ are n early constant. If $c_i(n)$ is nearly equal to C_i, then the solution to

$$A_n = C_1A_{n-1} + C_2A_{n-2} + \cdots + C_kA_{n-k} \qquad (**)$$

with initial conditions should be reasonably close to the sequence a_n. What can be said about the solution to the last equation? Without initial conditions not much can be said with certanty, but the following is usually true:
Principle 1. Let r be the largest root of the equation

$$r^k = C_1 r^{k-1} + C_2 r^{k-2} + \cdots + C_k r^0.$$

If this root occurs with multiplicity m, then there is usually a constant A such that solution to $(**)$ that satisfies our unspecified initial conditions is asymptotic to $An^{m-1}r^n$.

This principle gives much less accurate results than the use of generating functions or partial fractions, but generally requires much less work and gives an idea as to what to expect for $(*)$.
Principle 2. Suppose that $c_i(n) \to C_i$ as $n \to \infty$ and that at least one C_i is nonzero. Let r be the largest root of $(**)$. Then r^n is probably a fairly reasonable crude approximation to the solution a_n of $(*)$ that satisfy our unspecified initial conditions. Usually

$$\lim_{n\to\infty} (c_n)^{\frac{1}{n}} = r.$$

Example. Involutions.
Let a_n be the number of permutations of {1,2, ..., n} which are involutions, that is, no cycle lengths exceed two. Thus a recurrence relation for involutions is simple: either n is in a cycle with itself or it is in a cycle with one of the remaining $n-1$ elements. Thus

$$a_n = a_{n-1} + (n-1)a_{n-2},$$

with some appropriate initial conditions. The coeficients of this recursion are not almost constant, but there is a trick that works whenever coefficients are polynomials in n. Let $b_n = \frac{a_n}{(n!)^d}$, where d is to be determined.Dividing our recursion by $(n!)^d$, we obtain

$$b_n = \frac{1}{n^d} b_{n-1} + \frac{n-1}{(n^2-n)^d} b_{n-2}.$$

Now, if $d < \frac{1}{2}$, the last coefficient is unbounded, if $d > \frac{1}{2}$, both coefficients on the right approach zero. With $d = \frac{1}{2}$, the first coefficient approaches zero,

while the second approaches 1. Therefore, we are led to consider the recursion $b_n = b_{n-2}$ and hence the roots of the polynomial $r^2 = 1$. Since the largest root is 1, we expect b_n to approach 1. Thus, $(n!)^{\frac{1}{2}}$ is a rough approximation to a_n. Factorial can be eliminated by using Stirling's formula. Since our approximation is so crude, we can ignore factors like $\sqrt{2\pi n}$ and simply say that a_n probably grows roughly like $(\frac{n}{e})^{\frac{n}{2}}$.

3 Tools of the trade - continued

The theory of algorithms has undergone an extraordinarily vigorous development in the last 20 years, and probability theory has emerged as one of its most vital partners. The aim of this section is to provide basic tools from probability theory, specialy those frequently used in the analysis of algorithms. The selection in this section is by no means comperhensive, interested reader is refered to some of excellent introductions in probability theory [**?**], [**?**].

3.1 Notion of Probability

Even though we intuitively connect statements such as "this compendium will probably be a success" with probability, we will be concerned with much more idealized **model** of a particular conceptual experiment, with a well defined set of possible outcomes (**sample space**). We will **imagine** the experiment performed a great number of times. An **event** A is one of possible outcomes of the experiment. Then the meaning of the sentence *event A occurs with probability* 0.6 is that, in a long run, A is **expected** to appear sixty times out of a hundred. If A appears exactly k times out of n times the experiment has been performed, $\frac{k}{n}$ if called a relative frequency of the event A. Thus relative frequencies of events are expected to fluctuate about some fixed number that is called the probability of event.

Perhaps you saw this fraction $\frac{k}{n}$ before and took it as a definition of probability? It indeed coincides with probability in a special case when all events are equaly likely, i.e. when we have classical probability algebras. In the case of classical probability algebras , probabilities can be determined by combinatorial methods.

Example. A person having N keys in his pocket wishes to open his appartment. he takes one key after another from his pocket and triies to open the door. What is the probability that he will find the right key precisely at k-th trial?

Suppose that the $N!$ possible sequences of keys are all equaly likely (have the same probability). The answer is then extreemly simple: there are exactely $(N-1)!$ permutations in which certain key allways occupies the k-th position. The probability in question is therefore $\frac{(N-1)!}{N!} = \frac{1}{N}$. The situation, of course, is not so simple if keys are on a ring or can be tried more than once.

Example. An urn contains M red and $N-M$ white balls. Balls are drawn from the urn one after the other without replacement. What is the probability of obtaining the first red ball at the k-th drawing?
In order to answer the question, we must know what our sample space and events are. Since balls are drawn in order, we are speaking about permutations. We are interested in finding how many of $N!$ permutations have $k-1$ white balls, followed by a red ball at the k-th position. The first $k-1$ balls can be choosen in $\left(\begin{array}{c} N-M \\ k-1 \end{array} \right)$ ways from $N-M$ white balls; furthermore, these can be arranged in $(k-1)!$ ways. The red ball for k-th position can be choosen in M ways. The remaining positions can be arranged in $(N-k)!$ ways. Hence, the answer is

$$P_k = \frac{1}{N!} \left(\begin{array}{c} N-M \\ k-1 \end{array} \right) (k-1)!M(N-k)!.$$

(Note: previous example is a special case of this with $M = 1$.)

Example. An urn contains N balls, of which $M \geq 1$ are red and the rest is white. From the urn n balls are drawn. What is the probability of obtaining k red balls and $n-k$ white ones?
This question could have been worded differently. For example: in a serial production of machine parts, a series of N parts contains M rejects. What is the probability that b y taking a sample of n parts, this sample will contain k rejects? The point here is that balls and urns are just convinient model for great variety of 'real' situations we use probability in.
A sample of n elements out of N elements can be choosen in $\left(\begin{array}{c} N \\ n \end{array} \right)$ ways. Suppose that every such combination is equaly likely. Then the probability of every such combination is $(\left(\begin{array}{c} N \\ n \end{array} \right))^{-1}$. Therefore, we have only to count how many combinations contain k rejects. There are $\left(\begin{array}{c} M \\ k \end{array} \right)$ ways to choose k elements from m elements and $\left(\begin{array}{c} N-M \\ n-k \end{array} \right)$ ways for the rest. The answer is thus

$$P_k = \left(\begin{array}{c} M \\ k \end{array} \right) \left(\begin{array}{c} N-M \\ n-k \end{array} \right) (\left(\begin{array}{c} N \\ n \end{array} \right))^{-1}.$$

3.2 Kolmogorov Probability Spaces

So far we have given just the most intuitive notion of probability. It was restricted to finite number of possible events. We are now going to give a more formal definition of probability space, one that includes infinite number of events. This definition is due to Kolmogorov and is therefore often refered to as Kolmogorov probability theory.

We assume that algebra of sets is given, isomorphic to the algebra of events dealt with. (In algebra of sets we talk about unions and intersections, while in algebra of events about sums and products, i.e. the sum of two events $A + B$ is an event which occurs exactly when either A or B occurs, the product AB is an event which occurs only if A and B occur.) We assume further that this algebra contains not only the sum of two sets belonging to it but also the sum of denumerably many sets belonging to algebra. Such algebras are called σ-algebras or *Borel* algebras.

The following axioms are assumed in Kolmogorov theory:

- Let there be given a nonempty set Ω. The elements of Ω are said to be elementary events and are denoted by ω.
- Let $\mathcal{A}$ be an algebra of sets of the subsets of Ω; the sets A of $\mathcal{A}$ are called events.
- $\mathcal{A}$ is a σ-algebra, i.e.

$$A_k \in \mathcal{A} \quad (k = 1, 2, \ldots) \Rightarrow \sum_{k=1}^{\infty} A_k \in \mathcal{A}.$$

From the above axioms it follows immediately that if $A_k \in \mathcal{A}, k = 1, 2, \ldots$, then also $\prod_{k=1}^{\infty} A_k \in \mathcal{A}$. The next three axioms deal with properties of probabilities.

- To each element A of $\mathcal{A}$ is assigned a nonnegative real number $P(A)$ (sometimes p_A), called the probability of the event A.
- $P(\Omega) = 1$.
- If $A_1, \ldots, A_n, \ldots$ is a finite or denumerably infinite sequence of pairwise disjoint sets belonging to $\mathcal{A}$, then

$$P(A_1 + A_2 + \ldots + A_n + \ldots) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$$

The last axiom is called σ-additivity of the set function $P(A)$.

A σ-algebra $\mathcal{A}$ of subsets of a set Ω on which a set function $P(A)$ is defined such that all the above axioms are fullfiled will be called a **probability space in the sense of Kolmogorov** and will be denoted by $[\Omega, \mathcal{A}, P]$.

Every finite probability algebra is a Kolmogorov probability space, since an additive set function on a finite algebra of sets is trivially σ-additive.

Now, the last axiom is clearly not valid for events that are not mutualy disjoint. Even though the following theorem can be stated in a more general setting, we will use this simplest form:

Theorem 3.1 *For any two events A_1 and A_2 the probability that either A_1 or A_2 or both occur is given by*

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$

3.3 Conditional Probabilities

We introduced the notion of probability by means of relative frequences. Accordingly, in order to introduc e the notion of conditional probability, we examine conditional relative frequences first. If an event B occurs exactely n times in N trials and if among these n trials event A occurs k times together with event B, then the quotient $\frac{k}{n}$ is called the conditional frequency (denoted $f_{A|B}$) of A with respect to condition B. If f_B denotes relative frequency of event B in the whole sequence of trials ($f_B = \frac{n}{N}$), and f_{AB} denotes the relative frequency of both events appearing together ($f_{AB} = \frac{k}{N}$), then

$$f_{A|B} = \frac{f_{AB}}{f_B}.$$

Since f_{AB} fluctuates around $P(AB)$ and f_B fluctuates around $P(B)$, then $f_{A|B}$ fluctuates around

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad given \;\; P(B) > 0.$$

$P(A|B)$ is called the **conditional probability** of the event A with respect to the condition B.

Example. Four balls are placed successively into four cells, all 4^4 arrangements being equally probable. Given that the first two balls are in different

cells (event B), what is the probability that one cell contains exactly three balls (event A)? Given event B, it is easy to see that event A can occur in exactly two ways, and therefore $P(A|B) = \frac{2}{4^2}$. (It is easy to verify directly that events B and AB contain $4 \cdot 3 \cdot 4^2$ and $4 \cdot 3 \cdot 2$ points.)

Example. Consider families with exactly two children. Letting b and g stand for boy and girl respectively, first letter denoting the older child. There are four possibilities: bb, bg, gb, gg. Those are our sample points, each having probability $\frac{1}{4}$. Given that a family has a boy (event B), what is the probability that both children are boys (event A)? The event AB means bb, while B means bb, bg, gb. Therefore, the answer is $\frac{1}{3}$. However, most people expect the answer to be $\frac{1}{2}$. Why? $\frac{1}{2}$ is the correct answer to a similar question, namely: a boy is choosen at random and found to come from a family with two children; what is the probability that the second child is a male? Now, what is the difference between these two formulations of a question? The difference is that in the latter formulation the event B consist of only two sample points, that is bb and bg.

From the mathematical point of view the conditional probability can be considered as a new **probability measure**. Indeed, let Ω be an arbitrary set, $\mathcal{A}$ a σ algebra, P a probability measure (i.e. a nonnegative, completely additive set function satisfying $P(\Omega) = 1$), B a fixed element of $\mathcal{A}$ such that $P(B) > 0$. Then $P(A|B)$ is a probability measure on $\mathcal{A}$ as well, and $[\Omega, \mathcal{A}, \mathcal{P}(\mathcal{A}|\mathcal{B})]$ is again a Kolmogorov probability space.

3.4 The Independence of Events

Generally, the conditional probability $P(A|B)$ is different from $P(A)$. If, however, it is not, i.e. if $P(A|B) = P(A)$ then we say that A **is independent of** B. If A is independent of B, then B is also independent of A. Indeed from the definition of conditional probability it follows that

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

This equation and independence of A (of B) imediately implies that

$$P(B|A) = P(B).$$

Thus we can say that events A and B are independent of each other. It is easy to see that if A and B are independent,

$$P(AB) = P(A)P(B).$$

The complete independence of more than three events is defined similarly. The events $A_1, A_2, \ldots, A_n$ are said to be **completely independent** if for any $k = 2, 3, \ldots, n$ the relation

$$P(A_{i_1} A_{i_2} \ldots A_{i_k} = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k})$$

is valid for any combination $(i_1, i_2, \ldots i_k)$ from the numbers $1, 2, \ldots, n$. This equation represents $2^n - n - 1$ conditions (since any combination of k elements out of n can be choosen in $\left(\begin{array}{c} n \\ k \end{array} \right)$, then summing over all k gives 2^n but we have overcounted for $\left(\begin{array}{c} n \\ 0 \end{array} \right)$ and $\left(\begin{array}{c} n \\ 1 \end{array} \right)$, therefore we subtract the last two terms from the sum).

3.5 Probability Distributions

With respect to finite probability algebras, we can define a concept of **complete system of events** as follows: a system $A_1, A_2, \ldots, A_n$ is a complete system of events if the relations

$$A_i A_j = 0 \quad for \quad i \neq j \quad and \quad A_1 + A_2 + \ldots + A_n = \Omega$$

are valid, with all $A_i \neq 0$ for all $i = 1, 2, \ldots, n$.

This definition can be extended to the arbitrary probability space in the following manner: a finite or denumerably infinite system of events $\{A_n\}$ $(A_n \in \mathcal{A}, \setminus = \infty, \in, \ldots)$ is said to be complete (in the wider sense), if for $for \quad i \neq j \quad A_i A_j = 0$ and if occurence of the event A_n is almost sure i.e. if

$$P(\sum_n A_n) = \sum_n P(A_n) = 1.$$

Thus, this time instead of $A_1 + A_2 + \ldots = \Omega$ we only require that $P(\overline{\Omega'}) = 0$ holds, where $\Omega' = \sum_n A_n \subset \Omega$.

A sequence of probabilities of a complete system of events will be called a **probability distribution**.

4 Probabilistic Method

The probabilistic method has recently been developed intensively. One of the major reasons for this development is the increasingly important role of randomness in theoretical computer science. The basic probabilistic method can be described as follows: in order to prove the existence of a combinatorial structure with certain properties, we construct an appropriate probability space and show that a randomly chosen element in this space has the desired properties with positive probability. We will illustrate the method on a simple example.

Example. A Ramsey number $R(k,l)$ is the smallest integer n such that in any two-coloring of the edges of a complete graph on n vertices K_n by red and blue, there is either a red K_k (i.e. a complete subgraph on k vertices, all of whose edges are colored red) or a blue K_l. Ramsey (1939) showed that $R(k,l)$ is finite for any two integers k and l. Let us obtain a lower bound for the diagonal Ramsey numbers $R(k,k)$.

Theorem 4.1 *If* $\binom{n}{k} 2^{1-\binom{k}{2}} << 1$, *then* $R(k,k) > 2^{\frac{k}{2}}$ *for all* $k \geq 3$.

Proof. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equaly likely. For any fixed set R of k vertices, let A_R denote the event that the induced subgraph of K_n on R is monochromatic (all edges of the same color). Clearly, $Pr(A_R) = 2^{1-\binom{k}{2}}$. Since there are $\binom{n}{k}$ choices for R, the probability that at least one of the events occurs is at most $\binom{n}{k} 2^{1-\binom{k}{2}} << 1$. Thus, with positive probability, no event A_R occurs and there is a two coloring of K_n without monochromatic K_k, i.e. $R(k,k) \geq n$.

Note that if $k \geq 3$ and $n = \lfloor 2^{\frac{k}{2}} \rfloor$, then

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \frac{n^k}{2^{\frac{k^2}{2}}} < 1$$

and hence $R(k,k) > 2^{\frac{k}{2}}$ for all $k \geq 3$.

4.1 Random Graphs

We give the three most common models for random graphs. In all cases, graph G has n vertices.

- *Dynamic.* Imagine G to have no edges at time 0. At each time unit a randomly choosen edge is added to G. Then G evolves from empty to full.

- *Static.* Given e , let G be choosen randomly from among all graphs with e edges.

- *Probabilistic.* Given p, let the distribution of G be defined by $Pr[\{i,j\} \in G] = p$ for all i, j with these probabilities mutually independent (i.e. flip the coin, head with probability p, to determine if each edge is in G).

When $p = e(\begin{pmatrix} n \\ 2 \end{pmatrix})^{-1}$, the static and the probabilistic model are nearly identical. We will work mostly with probabilistic model. The notation that we use is $G(n,p)$.

Definition 4.1 *$r(n)$ is called a* **treshold function** *for a graph theoretic property A if*

1. *when $p(n) << r(n)$, $\lim_{n\to\infty} Pr[G(n,p) \vdash A] = 0$*

2. *when $p(n) >> r(n)$, $\lim_{n\to\infty} Pr[G(n,p) \vdash A] = 1$*

or vice versa.

Here are some examples: Property *A: graph is not planar* has threshold function $p(n) = \frac{1}{n}$. Property *A: graph is connected* has threshold function $p(n) = \frac{\ln n}{n}$. Property *A: graph contains a clique on k points* has threshold function $p(n) = n^{-2/(k-1)}$.

A bipartite graph $G = (U, V, E)$, $|U| = |V| = n$ is an (α, β)**-expander** if $|N(X)| \geq \alpha|X|$ for all $X \subset U$, $|X| \leq \beta|U|$ and the same holds for subsets of V.

Expanders are example of so called quasi-random graphs, i.e. graphs that have some properties of random graphs. They are, as we will see in [**?**] a good tool to use when an explicit construction of, for example, parallel sorting network, is desired.