Computation and Economics - Spring 2012
Section: Math re-fresher

Professor Sven Seuken
Department of Informatics, University of Zurich

Positioning: The ideas and concepts presented in this document are intended to provide an overview of the math used in the EconCS lecture and you will need to use them to complete your homework. We attempt to present it in an accessible and intuitive manner, so you can familiarize yourself with the concepts and keep them in mind when studying the course content. Most of the topics should already be familiar to you.

1 Basic notation

First we introduce some notation that will be valuable to keep in mind during the course.

Sets:

- A *set* is an unordered, possibly infinte collection of objects. The notation is
$$\{a, b, \ldots\},$$
where a is an *element* of the set. We can also write $a \in A$.
- A set B is a *subset* of A if it contains only elements that are also in A. We write
$$B \subseteq A.$$
- Operations on sets include *union* (combining all elements from 2 sets to a new set, denoted by $A \cup B$) and *intersection* (taking only the elements that are in both sets, denoted by $A \cap B$).
- $\emptyset$ denotes the empty set with no elements.
- $\mathcal{P}(A)$ denotes the *power set* of A, i.e. the set of all subsets of A. For example if $A = \{1, 2\}$, then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

Vector: A *vector* is an ordered, possibly infinte list of objects. The notation is
$$x = (x_1, x_2, \ldots),$$
where the x_i are *components* of the vector. [1]

[1] Note that we use the terms vector and *tupel* synonymously, even though conditions imposed on a vector are usually much more strict - let's ignore this for now.

Productspace: The *productspace* of 2 sets A and B is a set and denoted by

$$A \times B.$$

The *elements* of the productspace are all vectors of the form (a, b), where $a \in A$ and $b \in B$. For example: If 2 agents play a game where agent 1 will either get \$1 or nothing and agent 2 will get either \$2 or \$4. Then the sets of possible outcomes for each individual agent are $A_1 = \{0, 1\}$ and $A_2 = \{1, 2\}$. The outcomes of the overall game are given by $A_1 \times A_2 = \{(0, 2), (0, 4), (1, 2), (1, 4)\}$.

Mappings: A *mapping* f from a set A to a set B assigns an element of B to each element of A. Formally we write

$$f : A \to B.$$

If we want to say that f maps a specific $a \in A$ to $b \in B$, we write

$$a \overset{f}{\mapsto} b \text{ or } f(a) = b.$$

For example: The function $u(x) = 2x^2 - x^3$ is a mapping $u : \mathbb{R} \to \mathbb{R}$ and we have $u(0) = 0$ (or equivalently $0 \overset{u}{\mapsto} 0$).
Another interesting mapping is a *projection* that maps a vector on some componenten, e.g. $\pi_2(x_1, x_2, x_3) = x_2$ is a projection on the second component.[2]

2 Mathematical proofs

Given a precondition A say, we have to provide credible evidence that a consequence B holds. Here we informally present the 4 main proof rechniques and provide illustrative examples.

2.1 Proof by direct implication

Approach: We produce a series of logically correct implications of the form

$$A \Rightarrow A_1 \Rightarrow \ldots \Rightarrow A_n \Rightarrow B.$$

Example: Show that the product of two integers is even if one of them is even.

Precondition: One of two integers a, b is even. Without loss of generality let a be even, otherwise switch the roles of the numbers.

Implications:

$$\begin{aligned} &\Rightarrow \quad a = 2a' \text{ for some integer } a' \\ &\Rightarrow \quad ab = 2a'b \end{aligned}$$

Consequence: The product ab is again even.

[2]You will probably encouter more complex mappings in your studies: Consider taking the expectation of a random variable: Effectively this assigns a value to each element of a space of random variables. Note that the input to the mapping 'expectation' is itself a function. Mappings that assign a real number to a function are called *funcionales*.

2.2 Proof by complete enumeration

Approach: The problem is split into sub-problems, which can each be solved separately. Generally the split introduces additional preconditions that can be used in the sub-problems.

Example: Show that an agent in a sealed-bid second-price auction is best off bidding truthfully, i.e. bidding his true value.

Precondition:

- n agents have private values v_i for $i \in \{1, \dots, n\}$ for the subject.
- All agents have utility

$$u_i = \begin{cases} v_i - price, & \text{if the agent wins and pays } price, \\ 0, & \text{otherwise.} \end{cases}$$

- Each agent submits a bid b_i unknown to all other agents.
- If agent i with bid b_i is highest bidder, he wins the auction and pays a price equal to the second-highest bid, i.e. $price = \max\{b_j; j \neq i\}$.

Separation and proof: Let's consider agent 1 with private value v_1 and bid b_1. Let $b_{-1} := max\{b_j; j \neq 1\}$ be the maximum of all other bids.
First we separate the 3 cases: 1. $b_1 = v_1$, 2. $b_1 < v_1$, 3. $b_1 > v_1$.

1. If $b_1 = v_1$, agent 1 has utility

$$u_1^{=} = \begin{cases} 0, & \text{if } b_1 \leq b_{-1}, \\ v_1 - b_{-1}, & \text{if } b_{-1} < b_1, \end{cases}$$

and $v_1 - b_{-1} > 0$, since $v_1 = b_1 > b_{-1}$.

2. If $b_1 < v_1$, agent 1 has utility

$$u_1^{<} = \begin{cases} 0, & \text{if } b_1 < v_1 \leq b_{-1}, \\ 0, & \text{if } b_1 \leq b_{-1} < v_1, \\ v_1 - b_{-1}, & \text{if } b_{-1} < b_1 < v_1. \end{cases}$$

Note that by bidding lower than v_1, agent 1 receives 0 (instead of a positive utility $v_1 - b_{-1}$) in the second case and the same otherwise.

3. If $b_1 > v_1$, agent 1 has utility

$$u_1^{>} = \begin{cases} 0, & \text{if } v_1 < b_1 \leq b_{-1}, \\ v_1 - b_{-1}, & \text{if } v_1 < b_{-1} < b_1, \\ v_1 - b_{-1}, & \text{if } b_{-1} \leq v_1 < b_1. \end{cases}$$

Note that by bidding higher than v_1, agent 1 receives a negative utility $v_1 - b_{-1}$ (instead of 0 when bidding $b_1 = v_1$) in the second case and the same otherwise.

We see that $u_1^{=}$ is at least as high as both $u_1^{<}$ and $u_1^{>}$, and in some cases higher.

Consequence: Bidding truthfully in a sealed-bid second-price auction is at least as good as any other strategy. (In this case truthful bidding is called a *weakly dominant* strategy.)

2.3 Proof by contradiction

Approach: We show that a consequence holds by showing that the contrary cannot be true. This exploits the fact that saying $A \implies B$ is equivalent to saying $\neg B \Rightarrow \neg A$.

Example: Show that there are infinitely many prime numbers.

Assumption (denoted "$\mathbb{A}''$): There exists a largest prime number.

Implication: Then we can write a list $p_1, \ldots, p_n$ of all prime numbers and define $d = p_1 \cdot \ldots \cdot p_n + 1$. It is clear that no prime number divides d evenly, because by construction the remainder is always 1.
d is larger than the largest prime number, hence it cannot be a prime number itself. Then there must exist a number $d' < d$ that divides d.
If d' is not a prime number, we find a $d'' < d'$ that divides d' and hence also divides d.
We repeat finding smaller divisors, and because d is finite we must eventually find a prime number p_i that divides d.
But this contradicts the fact that no prime number divides d.

Consequence: Because the assumption that there are only finitely many prime numbers leads to a contradiction of a know fact, the assumption cannot be correct. Hence the opposite must be true.

2.4 Proof by induction

Approach: Use induction to show a consequence holds for all natural numbers $n = 0, 1, 2, \ldots$. Proceed as follows:

- Show: The consequence holds for $n = 0$.
- Show: If the consequence holds for an $n \geq 0$, then it holds for $n + 1$ as well.

Example: Prove the formula for the geometric sum: For any $n = 0, 1, 2, \ldots$ and any $\delta \neq 1$

$$\sum_{k=0}^{n} \delta^k = \frac{1 - \delta^{n+1}}{1 - \delta}.$$

Base clause: Clearly $\sum_{k=0}^{0} \delta^k = \delta^0 = 1 = \frac{1-\delta}{1-\delta}$.

Induction hypothesis: $\sum_{k=0}^{N} \delta^k = \frac{1-\delta^{N+1}}{1-\delta}$ holds for some fixed N.

Induction step: We need to show that if the induction hypothesis holds for N, than the consequence holds for $N + 1$.
For $N + 1$ we have

$$\sum_{k=0}^{N+1} \delta^k = \sum_{k=0}^{N} \delta^k + \delta^{N+1}.$$

We use the induction hypothesis to substitute the sum and get

$$\begin{aligned} \sum_{k=0}^{N+1} \delta^k &= \frac{1 - \delta^{N+1}}{1 - \delta} + \delta^{N+1} = \frac{1 - \delta^{N+1}}{1 - \delta} + \frac{(1 - \delta)\delta^{N+1}}{1 - \delta} \\ &= \frac{1 - \delta^{N+1} + (1 - \delta)\delta^{N+1}}{1 - \delta} = \frac{1 - \delta^{N+1} + \delta^{N+1} - \delta^{N+2}}{1 - \delta} = \frac{1 - \delta^{(N+1)+1}}{1 - \delta}. \end{aligned}$$

Consequence: The equation holds for $n = 0$. We were able to show that if it holds for an $n \geq 0$, then it also holds for the next natural number. Thus it must hold for $n = 1$, thus for $n = 2$, and so forth. This proves the equation for all natural numbers.

3 Calculus and series

3.1 Derivatives

Intuition: For a function f the derivative f' is a mapping that assigns the slope of the function f to each point.[3] The second derivative is the derivative of f'.

Product rule: $(fg)'(y) = f'(y)g(y) + f(y)g'(y)$.

Quotient rule: $\left(\frac{f}{g}\right)'(y) = \frac{f'(y)g(y) - f(y)g'(y)}{g^2(y)}$.

Chain rule: $(f \circ g)'(y) = f'(g(y))g'(y)$.

Derivatives of polynomials: $(f + g)'(y) = f'(y) + g'(y)$,$(af)'(y) = af'(y)$ and if $f(y) = ay^k$, then $f'(y) = aky^{k-1}$.

Exponential and Logarithm: $(\exp(\cdot))'(y) = \exp(y)$ and $(\log(\cdot))'(y) = (\cdot)^{-1}$.

Example: Let $f(x) := x \exp\left(x^2\right)$ where we would like the derivative. To get this we proceed as follows:

- Use the product rule to write
$$f'(x) = (x)' \exp(x^2) + x \left(\exp(x^2)\right)' = \exp(x^2) + x \left(\exp(x^2)\right)'.$$
- Use the chain rule to get the derivative of the last term and write
$$f'(x) = \exp(x^2) + x \exp'(x^2)(x^2)'.$$
- Use the polynomial and exponential rule to get
$$f'(x) = \exp(x^2) + x \exp(x^2)(2x) = (1 + 2x^2) \exp(x^2).$$

3.2 Finding extreme values

Suppose we have a function that maps our decision on a parameter to a certain outcome, e.g. a utility or a payoff. Let $f : \mathbb{R} \to \mathbb{R}$ denote this function. We are interested in the choice of the parameter that maximises the outcome. If f satisfies certain regularity conditions, i.e. it is twice continuously differentiable, then y is a local maximum[4] of f if and only if

$$f'(y) = 0 \text{ and } f''(y) < 0,$$

where f' and f'' denote the first and second derivative of f, respectively.

[3] Formally, the derivative of f in y can be defined by the limit $f'(y) := \lim_{h \to 0} \frac{f(y) - f(y+h)}{h}$, given the limit exists

[4] A local maximum is defined as follows: y is a local mximum if there exists an interval $[y - \epsilon, y + \epsilon]$ such that $f(y) \geq f(x)$ for all x in the interval. This is in contrast to a global maximum, where the value of f in the global maximum has to be greater than or equal to the value of f in any other point.

Example: An agent likes icecream and derives utility $u(x) = 2x^2 - x^3$ from eating $x$kg of icecream (first he really enjoys having more, but eating too much makes him sick). If icecream was free, how much should the agent consume?
We take the derivative $u'(x) = 4x - 3x^2 = x(4 - 3x)$ and set it equal to 0. This is the case for $x_1 = 0$ and $x_2 = \frac{4}{3}$, which provides the candidates for local maxima.
The second derivative is $u''(x) = 4 - 6x$, and evaluated at x_1 and x_2 we get $u''(x_1) = 4 > 0$ and $u''(x_2) = -8 < 0$.
This means that u has a local maximum[5] at $x_2 = \frac{4}{3}$, but not at $x_1 = 0$. x_1 is in fact a local minimum.
In a second exercise you could check what happens if the price for icecream is set to 50 cent/kg and utility is measured in \$.

3.3 Series

Introduction: A student is riding a bike towards a non-elastic concrete wall at a constant speed of 8m/s. The wall is 8m away. Secondly, a fly is traveling back and forth between the bike's front tyre and the wall at 24m/s. How far will the fly have travelled before the student hits the wall (resulting in the fly being squished between the wall and the front tyre of the bike)?
To answer this question we have 2 approaches:

- In the time it takes the student to get half way (a distance of 4m), the fly travels 12m, reaching the wall and coming back to meter 4. There the fly meets the student again. On its second flight to the wall, the fly reaches the wall and makes it back to meter 2 before meeting the student again. On every consecutive flight the fly gets half as far as on the previous flight. We get the distance of flight for the fly as an infinite sum

$$d_{fly} = 12 + 6 + 3 + \ldots = 12 \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k .$$

- Now note that the fly is 3 times as fast as the student. In the time it takes the student to get to the wall is 1 second. In this time the fly makes a distance of 24 meters.

If the first approach is valid, then we have found a value for the infinte sum: It should be 24![6]

Examples:

- Geometric series: For $|\delta| < 1$ we have

$$\sum_{k=0}^{\infty} \delta^k = \frac{1}{1-\delta}.$$

This makes sense if we consider the example for proof-by-induction: For large n the term δ^{n+1} in the numerator becomes very small, and hence the numerator converges to 1.

[5] Note that the global maximum and minimum of u on $\mathbb{R}$ are $\mp\infty$, respectively. If we restrict our attention however to a 'reasonable' interval like $[0, 2]$, x_2 becomes the global maximum.

[6] Semi-formal definition of convergence of series: Let $(a_k)_{k\geq 0}$ be a sequence of real numbers. We say that the series $\sum_{k=0}^{\infty} a_k$ *converges* if the sequence of partial sum $\left(\sum_{k=0}^{N} a_k\right)_{N\geq 0}$ converges. Note that proving convergence of series is not a trivial problem. The results presented here are subject to strict conditions which we will boldly ignore for the sake of bravity of the session.

- Harmonic series: The series $\sum_{k=0}^{\infty} \frac{1}{k}$ does not converge. Instead the partial sums can become arbitrarily large.
- Derivative of geometric series: Consider a function defined by $f(\delta) := \sum_{k=0}^{\infty} \delta^k = \frac{1}{1-\delta}$ for suitable values of δ. If we take the deivative on both sides (and boldly assuming that the equality survives), then we get $f'(\delta) = \sum_{k=0}^{\infty} k\delta^{k-1} = \frac{1}{\delta}\sum_{k=0}^{\infty} k\delta^k = \frac{1}{(1-\delta)^2}$. Multiplying by δ, we get the series representation

$$\sum_{k=0}^{\infty} k\delta^k = \frac{\delta}{(1-\delta)^2}.$$

- Discounting: Consider a game that provides the agent a fixed payoff of v in every round. There are two ways to think about this:
  - There is a probability $0 < 1-\delta < 1$ that the game will end after each round and the agent has no preference on the timing of payoffs, i.e. he is indifferent wether a payoff occurs now or in the future and only cares about the actual amount.
  - The game continues forever, but the agent has a discount for deferred payoffs, i.e. getting v units of utility one round later is only worth δv to the agent, getting it in round 2 is worth $\delta^2 v$, etc. Here δ is called the *discount factor*.

  We would like to show that the payoff $\sum_{k=0}^{\infty} \delta^k v = \frac{v}{1-\delta}$, the total discounted payoff to an agent facing an infinite sequence of payoff v with discount factor $0 < \delta < 1$, can be alternatively explained as the expected total payoff to an agent for whom $1-\delta$ is the probability the game will end in any round without discounting.
  Consider an agent playing a game that yields utility of v each round. At the end of each round it is decided with probability $1-\delta > 0$ that the game will stop, otherwise it continues at least one more round. The utility to the agent is the total accumulated utility over all rounds played. Hence the expected total utility is

$$\begin{aligned} E[u] &= \sum_{k=1}^{\infty} kvP[\text{Game ends after round k}] = \sum_{k=1}^{\infty} kv\delta^{k-1}(1-\delta) \\ &= v\left(\sum_{k=0}^{\infty} k\delta^{k-1} - \sum_{k=0}^{\infty} k\delta^k\right) = v\left(\frac{1}{\delta}\sum_{k=0}^{\infty} k\delta^k - \sum_{k=0}^{\infty} k\delta^k\right) \\ &= v\left(\frac{1}{\delta} - 1\right)\sum_{k=0}^{\infty} k\delta^k = v\frac{1-\delta}{\delta}\frac{\delta}{(1-\delta)^2} = \frac{v}{1-\delta}, \end{aligned}$$

  where we used the geometric series formula and it's 'derivative' as described previously.

4 Probability theory

4.1 Probability space

Definition: A *(discrete) probability space* is a tupel $(\Omega, \mathcal{A}, P)$, where

- Ω is a set, called the *sample space*. For example, this could be a finite set of actions avaiable to an agent $A = \{a_1, \ldots, a_n\}$, the set of natural numbers $\mathbb{N}$, or real numbers $\mathbb{R}$. The elements $\omega \in \Omega$ are called *elementary events*.

- $\mathcal{A}$ is a set of subsets of Ω.[7] You can assume for now that $\mathcal{A} = \mathcal{P}(A)$ is the power-set of Ω, i.e. the set of all subsets of Ω.
- The elements of $\mathcal{A}$ are the *events*. For instance consider a probability space that models tomorrow's weather. The event 'it rains' would be a set containing all elemenatry events for which the number of raindrops is not 0, and it would exclude any event for which the number is 0.
- The *probability distribution* is a mapping $P : \mathcal{A} \to [0,1]$ that assigns a *probability* to each event and statsifies some regularity conditions.[8]

Important attributes: A the probability distribution has the following key attributes:

- $P(\Omega) = 1$: probability of anything happening is 1.
- $P(\emptyset) = 0$: probability of nothing happening is 0.
- Let $A, B \in \mathcal{A}$ be events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. If A and B are disjoint (i.e. they have no common elements), then $P(A \cap B) = 0$ and hence $P(A \cup B) = P(A) + P(B)$. This is called *additivity.* [9]
- Let $A \subseteq B \subseteq \Omega$ be events, where A is a subset of B. This means that if A occurs, then B also occurs. These could be 'it rains all day long' and 'it rains at some point during the day'. Then $P(B \backslash A) = P(B) - P(A)$.

Examples:

- Consider the toss of a fair coin. The probability space is defined by
$\Omega = \{'heads', 'tails'\}$,
$\mathcal{A} = \{\{\}, \{'heads'\}, \{'tails'\}, \{'heads', 'tails'\}\}$,
$P('heads') = P('tails') = \frac{1}{2}$.
- Consider a loaded dice[10] thrown once. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. The event 'Even' could be defined as the subset $\{2, 4, 6\} \subseteq \Omega$. The probability for 'Even' can be calculated by $P('Even') = P(\{2\}) + P(\{4\}) + P(\{6\})$.[11]
- One important distribution is the uniform distribution on an interval of the real numbers. Consider $\Omega = [0, 1]$ with $P([a, b]) = b - a$ for any $0 \leq a \leq b \leq 1$. This is called a *uniform distribution*, because it assigns the same probability to all sub-intervals of the same length.
An important application is to randomly select a strategy: An agent will cooperate with probability p and defect with probability $1 - p$. Then p itself could be drawn uniformly from $[0, 1]$.

[7] In general, $\mathcal{A}$ must be a σ-algebra, satisfying a set of non-trivial conditions.

[8] On any probability space, a mapping P is a probability measure if the following holds: (I) $P(\Omega) = 1$ (probability of anything happening is 1) and (II) let $(A_i)_{i \geq 0}$ be a sequence of pairwise disjoint events, then $P(\bigcup_{i \geq 0} A_i) = \sum_{i \geq 0} P(A_i)$. We say that P is *σ-additive.*

[9] Intuitively if we want to know the probability that the weather will be bad tomorrow, we can look at the events 'rain' and 'storm' and add up the respective probabilities. But then we double-counted the possibility of simulaneous rain and storm and need to take out the probability of both things occuring at the same time once.

[10] Loaded dice are dice where (with or without intent) the numbers are not equally likely to occur.

[11] Note that if Ω is finite it is sufficient to define the probability distribution only on the elementary events.

4.2 Random variables

Definition: A ramdom variable is a mapping $X : \Omega \to I$ from a sample space Ω into an image space I.

- A common example for the image space is $\mathbb{R}$. In this case X assigns a value (e.g. a utility) to a certain even (e.g. 'Even').
- We write $P[X \in A] := P(\{\omega \in \Omega : X(\omega) \in A\})$ for short.
- The *expectation of* X is defined as $E(X) := \sum_{\omega \in \Omega} X(\omega)$. This is equivalent to writing $E[X] = \sum_{x \in I} xP[X = x]$.

Examples:

- Consider the loaded dice. You are playing a game where you gain \$1 if the result is even and loose \$1 otherwise. Then your payoff is given by a random variable

$$Y(\omega) = \begin{cases} 1, & \omega \in \{2,4,6\}, \\ -1, & \omega \in \{1,3,5\}. \end{cases}$$

  and the expectation is $E[Y] = 1 \cdot P('even') + (-1) \cdot P('odd')$.

- Consider the repeated toss of a fair coin and let Z denote the number of times the coin is tossed until it shows 'tails' for the first time. First we construct the probability space:

$$\Omega = \{(\omega_1, \omega_2, \ldots); \omega_i \in \{'heads', 'tails'\} \text{for all } i \geq 1\}.^{12}$$

  We can write Z as

$$Z(\omega) = \begin{cases} 1, & \text{if } \omega_1 =' tails', \\ 2, & \text{if } \omega_1 =' heads', \omega_2 =' tails', \\ 3, & \text{if } \omega_1 =' heads', \omega_2 =' heads', \omega_3 =' tails', \\ \ldots \end{cases}.$$

  This gives us a probability distribution for Z defined by $P[Z = k] = \left(\frac{1}{2}\right)^k$ for all $k \geq 1$. The expectation of X is $E[Z] = \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k = \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} = 2$, using the derivative of geometric series formula. Hence on average, we wait 2 rounds until 'tails' comes up.

4.3 Conditional probability, conditional expectation, Bayes' Rule

Definition: Let $(\Omega, \mathcal{A}, P)$ be a probability space and $A \in \mathcal{A}$ an event with probability $P(A) > 0$. Let $X : \Omega \to I$ be a random variable on this probability space.

- Then for any event $B \in \mathcal{A}$ we define the *conditional probability* of B given A by $P(B|A) := \frac{P(A \cap B)}{P(A)}$.
  Intuition: We may have an initial belief about the probability of B occuring. However, before knowing the exact elementary event that occurs, we know that it will be in A. We use this information *to update our belief* about the probability of B occuring.

[12] Note the case of the infinitely repeated coin toss, the probability of any elementary event is 0, because it is the product of infinitely many times $\frac{1}{2}$.

- We define the *conditional expectation* of X by $E[X|A] := \sum_{x\in I} xP[X = x|A]$, where $P[X = x|A] = P(\{\omega \in \Omega; X(\omega) = x\}|A)$.
  Intuition: Again, this is an update of our best belief about the expecation of X, given the information that the event A will occur.

Examples:
- In the loaded dice example, we could use A ='Even'. For B ='Odd' we get $P(B|A) = \frac{P(\{2,4,6\}\cap\{1,3,5\})}{P(\{2,4,6\})} = 0$, because $\{2, 4, 6\}$ and $\{1, 3, 5\}$ are disjoint.
  Recall that Y was a random variable, representing the payoff to you. The expected payoff, conditioned on A becomes
  $$E[Y|A] = 1 \cdot \frac{P(A \cap A)}{P(A)} + (-1) \cdot \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} - \frac{P(\emptyset)}{P(A)} = 1.$$
  This means that even though we don't know the exact distribution, but somehow gained the information the the result will lay in A, then we can use this information to update our belief about the expected payoff from the game (and hence know we should definately play).
- Consider the repeated coin toss example and the event $A = \{Z \geq 2\}$. This is like saying that the first throw was not 'tails'. We calculate the conditional expectation:
  $$\begin{aligned} E[Z|A] &= \sum_{k=1}^{\infty} kP[Z = k|Z \geq 2] = \sum_{k=1}^{\infty} k\frac{P[Z = k \text{ and } Z \geq 2]}{P([Z \geq 2]} \\ &= \left(\sum_{k=1}^{\infty} kP[Z = k] - 1 \cdot P[X = 1]\right)\left(\frac{1}{P[Z \geq 2]}\right) \\ &= (E[Z] - P[Z = 1])\left(\frac{1}{\frac{1}{2}}\right) = \left(2 - \frac{1}{2}\right) \cdot 2 = 3. \end{aligned}$$
  Note that once the first throw was not 'tails' this just pushes the expected waiting time up by 1.

Bayes' Rule: In addition to the assumptions made above, let $P(B) > 0$, such that both $P(A|B)$ and $P(B|A)$ are welldefined. Then $P(A|B)$ can be expressed as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$[13]

An extended form of Bayes' Rule is often applied, which makes use of the fact that $P(B) = P(B|A)P(A)+P(B|A^C)P(A^C)$. Here A^C denotes the complement $A^C := \Omega\backslash A$ of A. Plugging this into the equation we get

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

Example: Assume a patient has a certain illness with probability 0.1% (1 of 1000). A test has been developed that has the following attributes:

[13]Proof: We have $\frac{P(A\cap B)}{P(B)} = P(A|B)$ and $\frac{P(A\cap B)}{P(A)} = P(B|A)$. Therefore $P(A|B)P(B) = P(A\cap B) = P(B|A)P(A)$. Dividing by $P(B)$ gives the result.

- Used on a patient, who actually has the illness, the test will report a postive result with 99% accuracy (with 99 out of 100 affected patients the test detects the illness).
- Used on a patient, who does not have the illness, the test will report a negative result with 98% accuracy, but in 2% of the cases repots a false positive result.

This may seem like a decent test. But consider the following application of Bayes' rule: Let A denote the event that the patient is ill and let B denote the event that the test returns a positive result. From the previous information we get the following probabilities:

- $P(A) = 0.001$ for the patient having the illness,
- $P(B|A) = 0.99$ for the test correctly detecting the illness,
- $P(B|A^C) = 0.02$ for the test falsely detecting the illness,
- $P(A^C) = 0.999$ for the patient being healthy.

We are interested in $P(A|B)$, the probability that if the test result is positive, the patient actually has the illness. Applying the extended Bayes' Rule we get

$$P(A|B) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.02 \cdot 0.999} \approx 5\%.$$

Thus, when a patient takes the test and the test is positive and he asks the doctor wether he has the disease or not, the doctor will still answer that it is still very unlikely that the patient has the disease.

5 Complexity and O-notation

5.1 Algorithmic complexity

Consider the following alogrithm where the array A is of length N:

```
bubbleSort(Array A)
for (n=A.size; n>1; n=n-1)
  for (i=0; i<n-1; i=i+1)
    if (A[i] > A[i+1])
      A.swap(i, i+1)
```

We would like to know how many times the (potentially expensive) comparisons must be performed. Usually we look at the worst-case run-time and look for a upper bound:

- Outer loop: The outer loop runs exactly $N - 1$ times.
- Inner loop: The number of runs depends on the specific value of n and is $n - 1$ each time.

Taking this together we get

$$\sum_{n=2}^{N} \sum_{i=0}^{n-2} 1 = \sum_{n=2}^{N} (n-1) = \frac{N(N-1)}{2} = \frac{1}{2}N^2 - \frac{1}{2}N.$$

Thus, to sort an array of length N, the algorithm must perfom $\frac{1}{2}N^2 - \frac{1}{2}N$ comparisons.

5.2 O-notation

Consider runtime as a function of the input size. Usually it is inconvenient (and impossible in general) to describe the complexity of an algorithm by the exact number of steps. Instead we are more interested in the *asymptotic complexity*, which is given by the dominant behaviour of the runtime function: We say that a runtime function $f : \mathbb{N} \to \mathbb{N}$ is *of class* $O(g)$ for a function $g : \mathbb{N} \to \mathbb{N}$ if there exist constants $a, c \geq 0$ such that $f \leq ag + c$.

Bubble-sort: The runtime of bubble-sort is in $O(x^2)$. To see this, set $a = 1, c = 0$.
However the runtime is not in $O(x)$. The proof uses contradiction: Let $f(x) := \frac{1}{2}x^2 - \frac{1}{2}x$ denote the runtime of bubble-sort and assume that there exist $a, c \geq 0$ such that $f(x) \leq ax+c$ for all x. But then we would have $0 \leq -x^2 + (2a+1)x + c$ for all x. This is an upside-down parabola and hence negative for some sufficiently large x.[14]

Important complexities: We have that

- $O(x^{-1})$ is smaller than
- $O(1)$ *(constant)* is smaller than
- $O(\log(x))$ *(logarithmic)* is smaller than
- $O\left(x^{1/n}\right)$ is smaller than
- $O(x)$ *(linear)* is smaller than
- $O(x^n)$ *(polynomial)* is smaller than
- $O(2^x)$ *(exponential)*.

[14]You can verify that setting $x := (2a+q) + \frac{c}{2a+1}$ will do the job.