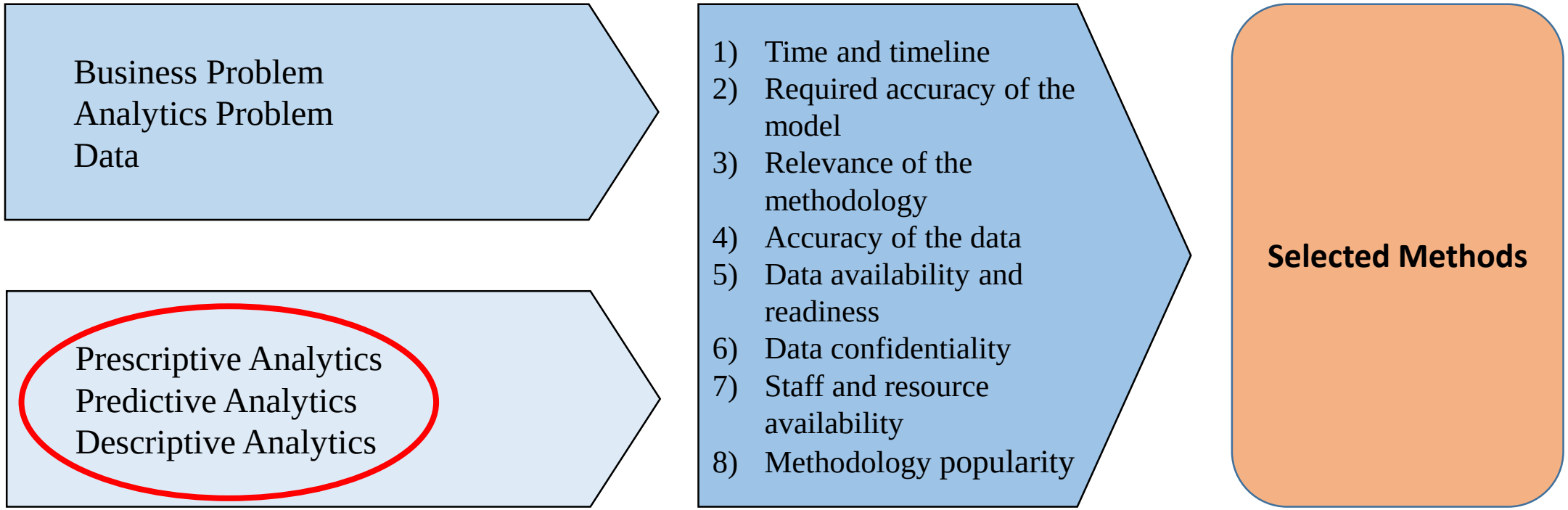
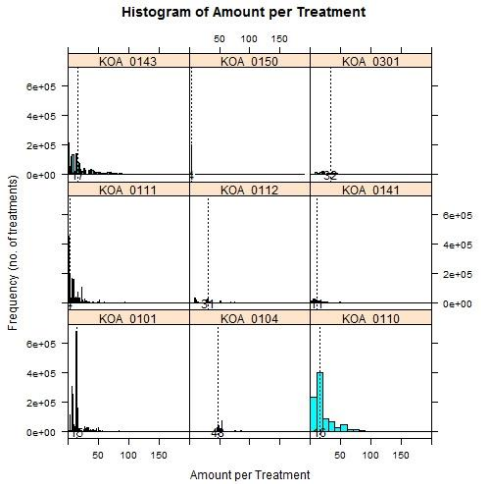
5.1 Identify available problem solving approaches / methods

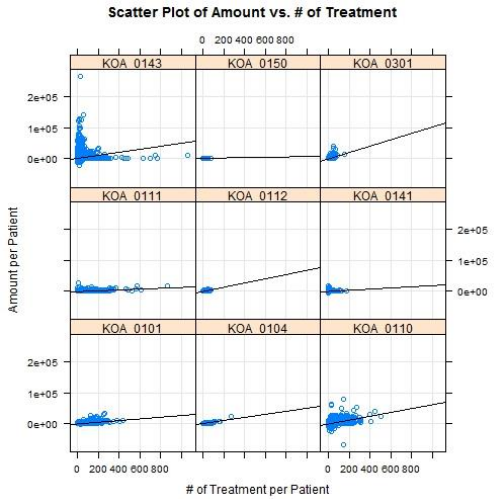


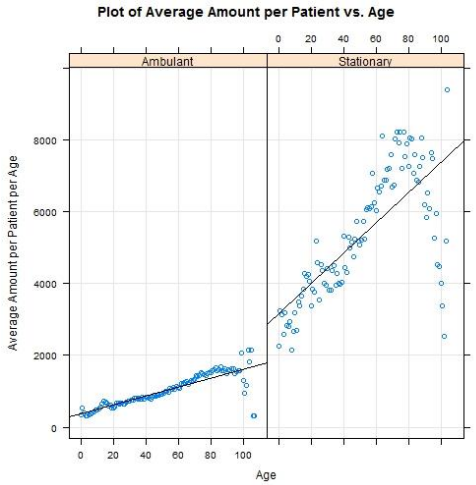
5.2 Descriptive Methods

Examples

- Mean, median, mode, variance, standard deviation, …
- Scatterplot, histogram, boxplot, steam-leaf plot





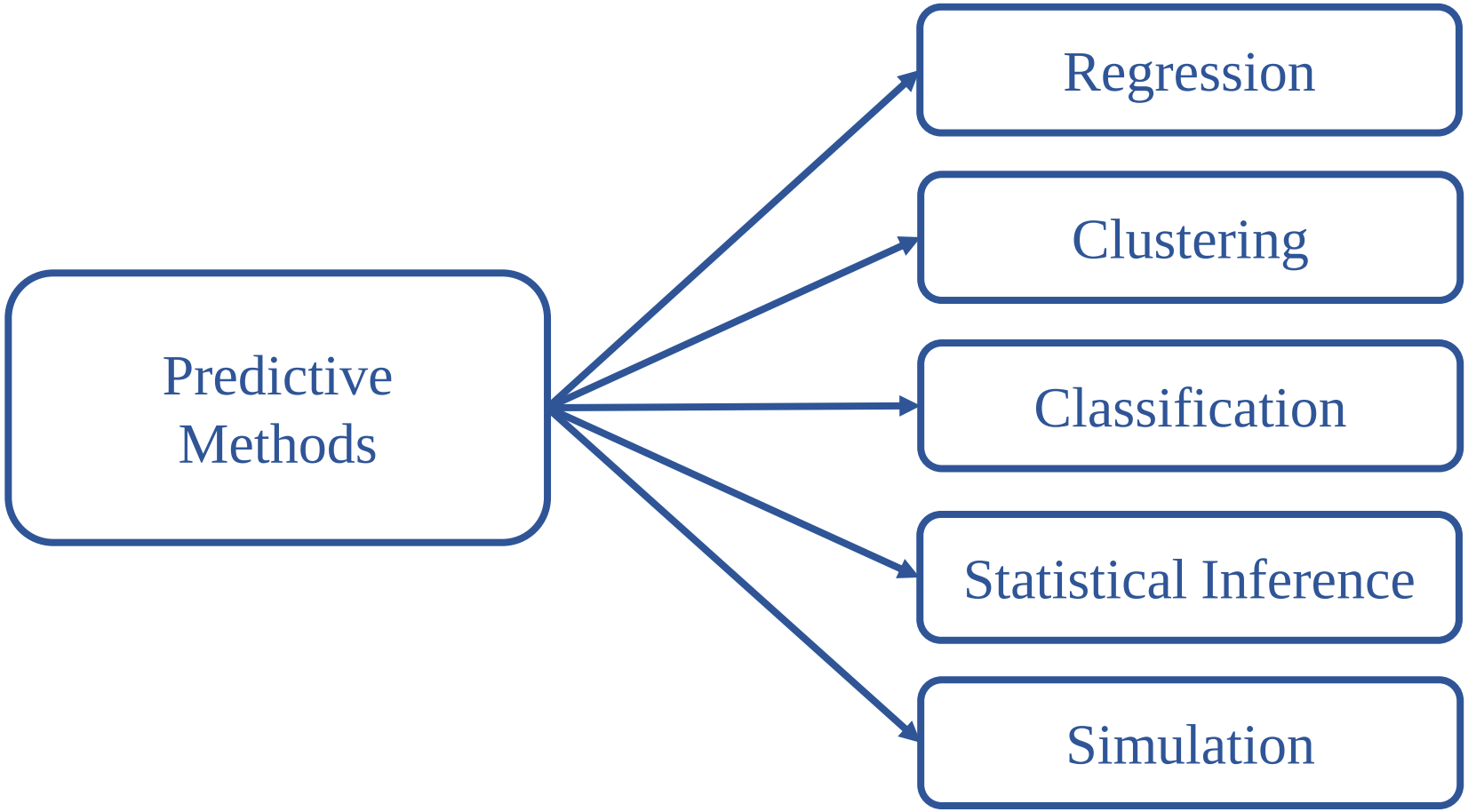


5.2 Descriptive Methods

When to use descriptive methods?

- For reducing, summarizing, and grouping data
- There is no dependent variable, i.e. one is simply describing what is or what the data shows
- Is typically used to simplify large amounts of data
- Is typically used if one is unfamiliar with the data set and needs to build up a “feeling” for the data

5.3 Predictive Methods



5.3 Predictive Methods

Regression

With **regression** we typically mean techniques for estimating relationships among variables and building up an understanding which variables are important in predicting future values.

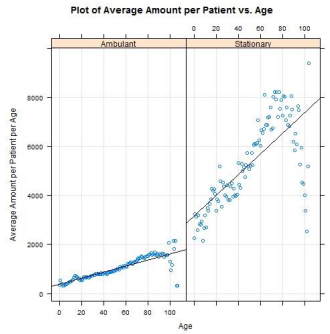
It consists of a dependent variable (the variable we want to estimate or to predict) and of independent variables (predictors).

5.3 Predictive Methods

Regression

Regression methods (examples):

Linear regression



Step-wise regression: Inclusion or deletion of the independent variable step by step based on some statistical measure like t-test or F-test:

- Forward selection: starting with no (independent) variable and adding then the variable which improves the model most by this inclusion.
- Backward deletion: start with all variables included and deleting then the variable which improves the model most by this deletion.

5.3 Predictive Methods

Regression

Regression methods (examples):

Shrinkage regression e.g. Ridge regression or Lasso

- If there are more variables than observations, least-square estimator does not exist
- If $(X^TX)^{-1}$ *) is singular or near singular, the least-square estimator does not exists

Thus, by introducing a penalty (additional constraints to the coefficients of X) one can overcome that issue.

*) $X^T = (x_1,\ldots,x_p)$, the independent variables

5.3 Predictive Methods

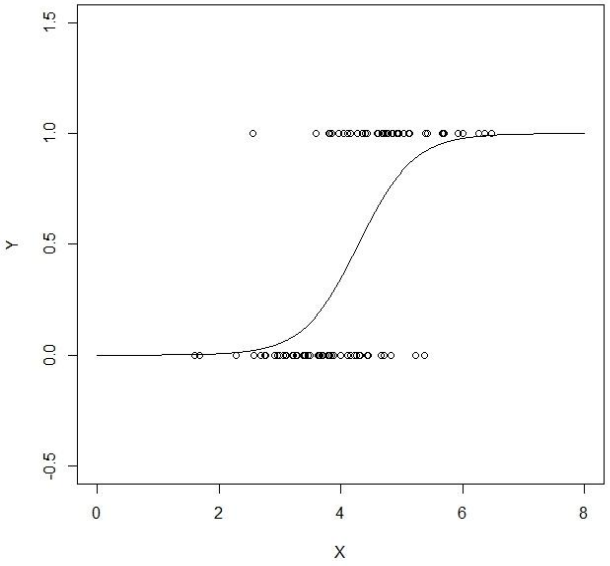
Regression

Regression methods (examples):

Logistic regression

The regression case where the dependent variable is categorical, especially binary i.e. has only two values e.g. 0 and 1.

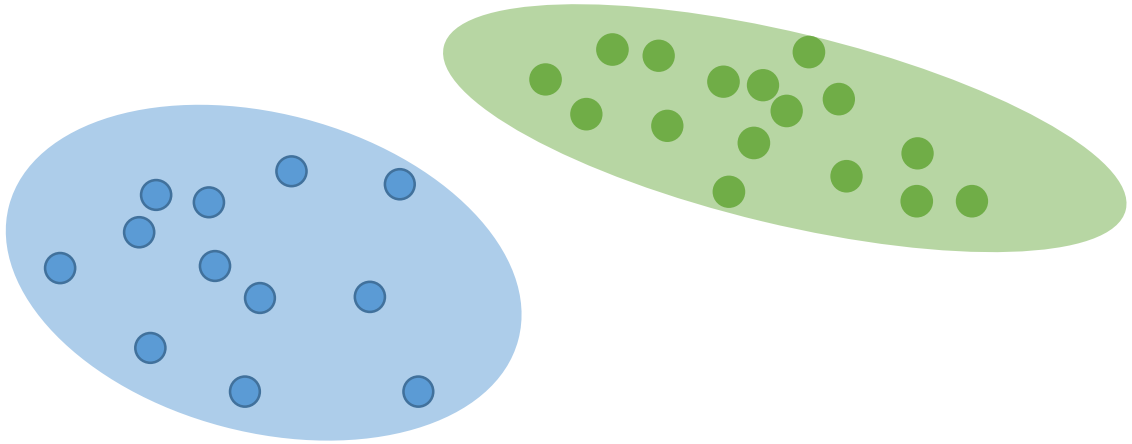
This model is important in credit scoring models and thus, in the financial risk model area very often used.



5.3 Predictive Methods

Clustering

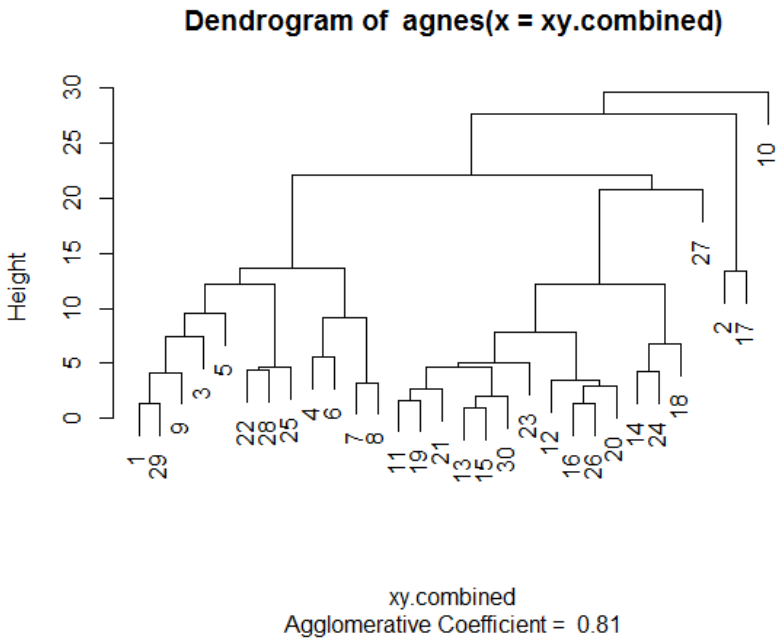
Clustering is a techniques for the segmentation of data into “naturally similar groups”.



5.3 Predictive Methods

Cluster methods (examples)

Hierarchical clustering: If one wants an ordered set of clusters with observation precision, i.e. based on the similarities of the observations or more precise a measure of dissimilarity i.e. a metric of distance between observations.



5.3 Predictive Methods

Clustering

Cluster methods (examples)

k-means clustering: If one wants to partition n observations into k clusters, where k is known.

It is a so-called centroid based method, i.e. it aims to minimize the distance (as the sum of squares) of each point in the cluster to the cluster center.

x-means clustering: If the number of clusters x is unknown

5.3 Predictive Methods

Classification

Classification methods are techniques for the prediction of the group membership of the observations.

…and there are many, many classification algorithms...

5.3 Predictive Methods

Classification

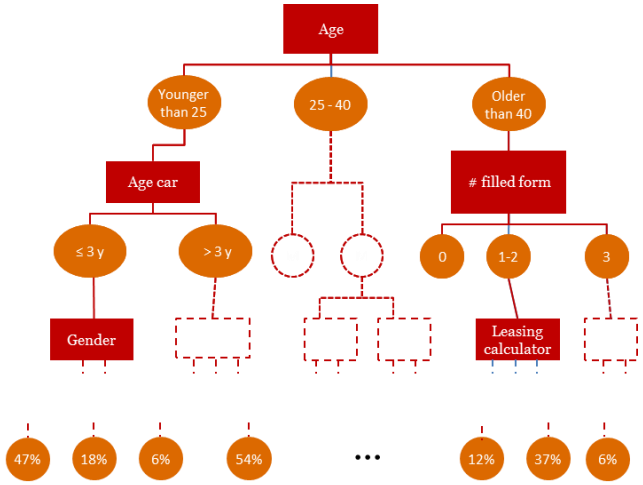
Classification methods (examples)

Decision trees: The algorithms are starting at the top and at each node a variable is chosen / determined that splits best the sample of observations.

Typically used when a transparent model is needed.

Many different types:

- CART
- MARS
- Random Forest
- Bagging



5.3 Predictive Methods

Classification

Classification methods (examples)

K-nearest neighbors: This is a non-parametric algorithm which classifies a point under consideration of the k-nearest neighbors to this point. One have first to train the algorithm on a training set such that it can memorize the characteristics.

It is typically used when the data dimension is not so high.

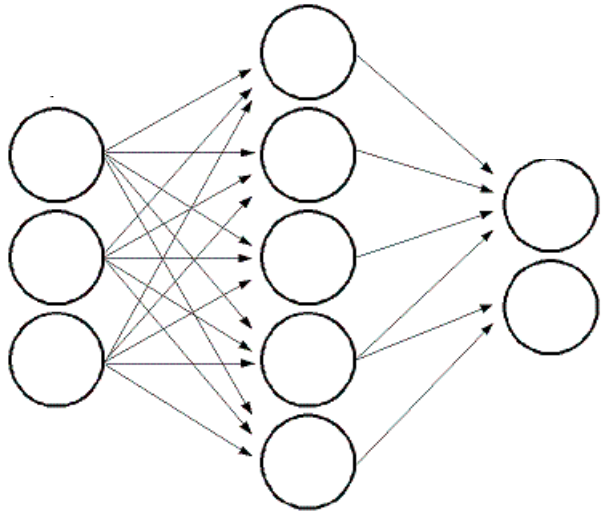
5.3 Predictive Methods

Classification

Classification methods (examples)

Neural Networks: The algorithms learns features in the data by changing the weights between the nodes based on learning rules. Thus, the algorithm also need first to be trained.

It is typically used if one have no idea about the features one is looking at or about the importance of a feature.



5.3 Predictive Methods

Classification

Classification methods (examples)

And if you do not know where to start: *Support Vector Machine (SVM)* or *Naïve Bayes*

5.3 Predictive Methods

Statistical Inference

Statistical inference means drawing conclusions based on data.

The most common known methods are

- Confidence intervals
- Hypotheses testing
- Analysis of variance (ANOVA)
- Design of experiments

5.3 Predictive Methods

Statistical Inference

Statistical inference (examples)

Design of experiments deals with the planning, conducting, analyzing and interpreting control tests. It aims to quantify the effects of the values of the output parameters by controlled variation of the input factors.

5.3 Predictive Methods

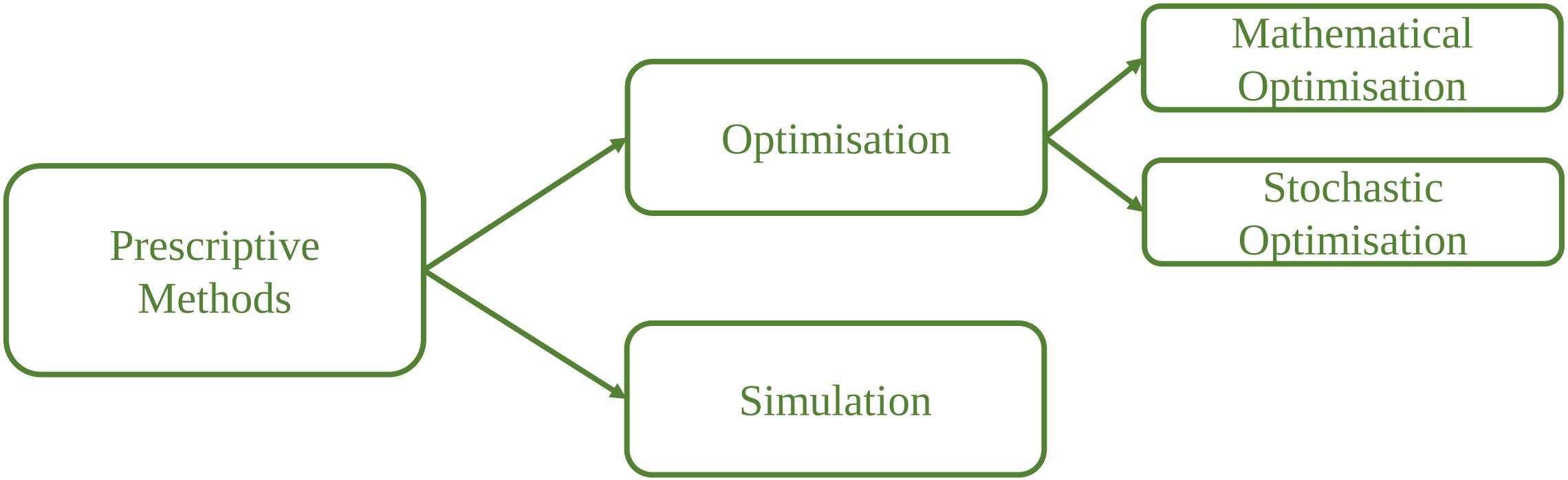
Simulation

Simulation is the design of a model of a real-world system or process, executing the model and analyzing the output of the model. The intention is “learning by doing”.

…and there are many, many types of simulations…

⇒ we will have a detailed look at them under prescriptive methods

5.4 Prescriptive Methods



5.4 Prescriptive Methods

Mathematical Optimisation

Optimization: Mathematical Optimization

Mathematical optimization problem:

minimize $f(x)$

subject to $f_i(x) \leq b_i$, $i = 1, \ldots, m$

Where

$x = (x_1, \ldots, x_n)$: optimization variables

$f : \mathbf{R}^n \rightarrow \mathbf{R}$: objective function

$f_i : \mathbf{R}^n \rightarrow \mathbf{R}$, $i = 1, \ldots, m$: constraint functions

Then, the optimal solution $x^\star$ has smallest value of f among all vectors that satisfy the constraints

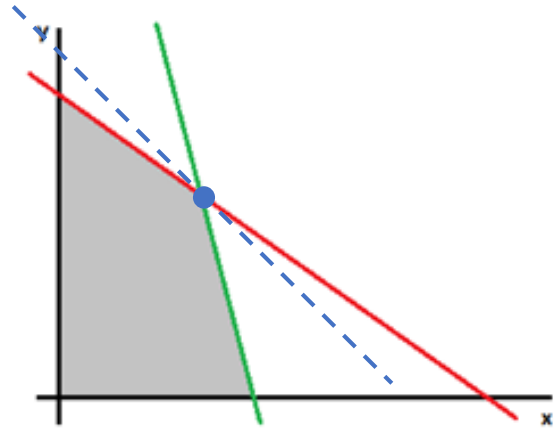
5.4 Prescriptive Methods

Mathematical Optimisation

Mathematical Optimization (examples)

Linear Programming determines the optimum in a linear mathematical model subject to linear equality or inequality constraints.

Integer Programming is a special case of linear programming where all variables are required to take on integer values only



5.4 Prescriptive Methods

Mathematical Optimisation

Mathematical Optimization (examples)

Nonlinear Programming determines the optimum of a function subject to equality and/or inequality constraints where either the function subject to optimization and/or the some of the constraints are non-linear.

Metaheuristics makes use of intensification and diversification. First, with randomized diversification diverse solutions are generated to explore the search space on a global scale. Then, with intensification a focus to the search in a local region is set by knowing that a current good solution is found in this region. It mixes a stochastic element with local search methods.

5.4 Prescriptive Methods

Mathematical Optimisation

Mathematical Optimization (examples) – but never used so far by us

Other techniques known from mathematics are *calculus of variations* e.g. Euler-Lagrange equations or the generalization which is known as *Optimal Control Theory*.

Dynamic Programming calculates the solutions of sub-problems and the overall solution can be derived out of the solutions of the sub-problems.

5.4 Prescriptive Methods

Stochastic Optimisation

Optimization: Stochastic Optimization

Minimizing the loss function $L = L(\theta)$:

$\Theta^* \equiv \arg\min_{\theta \in \Theta} L(\theta) = \{\theta^* \in \Theta : L(\theta^*) \leq L(\theta) \text{ for all } \theta \in \Theta\}$

where θ is the n-dimensional vector of parameters that are being adjusted and $\Theta \subseteq \mathbf{R}^n$.

5.4 Prescriptive Methods

Stochastic Optimisation

Optimization: Stochastic Optimization (cont'd)

Where either there is a random noise in the measurements of $L(\theta)$ i.e. $y(\theta) = L(\theta) + \varepsilon(\theta)$

and / or

there is a random choice made in the search direction of the iteration algorithm.

5.4 Prescriptive Methods

Stochastic Optimisation

Stochastic Optimization (examples)

Measurement with noise: *Stochastic Approximation* that is a recursive update rule.

The most famous algorithm is the Robbins-Monro Algorithm.

Random search: *Simulated Annealing*

5.4 Prescriptive Methods

Simulation

Repetition:

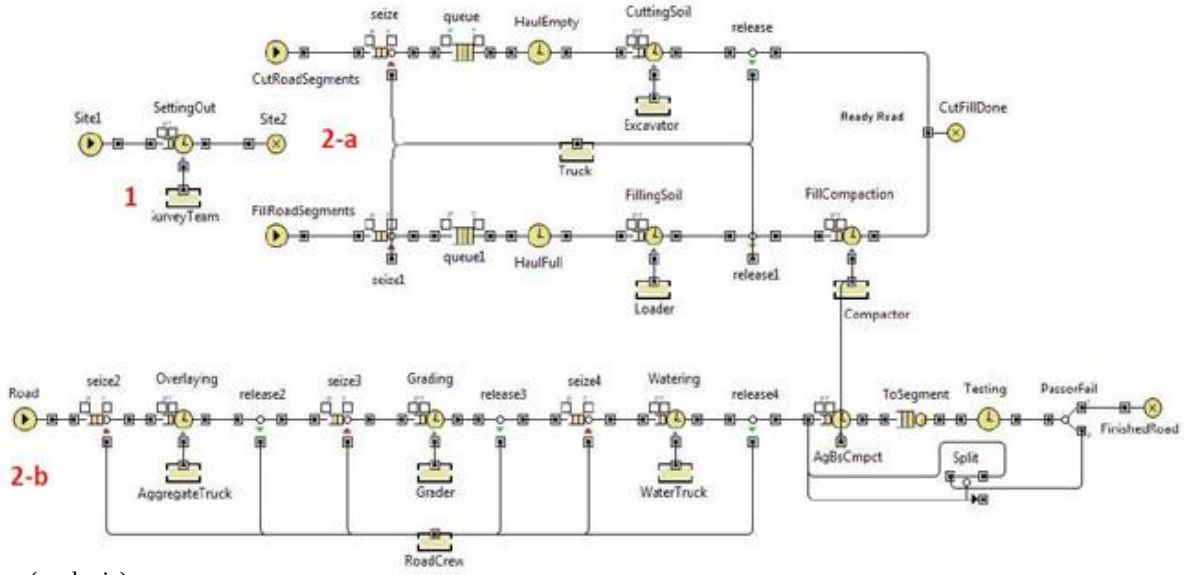
Simulation is the design of a model of a real-world system or process, executing the model and analyzing the output of the model. The intention is “learning by doing”.

5.4 Prescriptive Methods

Simulation (examples)

Discrete-Event Simulation simulates a discrete sequence of events where each event occurs at a particular instant in tim The model updates its state only at point in time when events occur. Between two time points i.e. two events there is no change in the system.

Example: Patients running through the operating room processes.



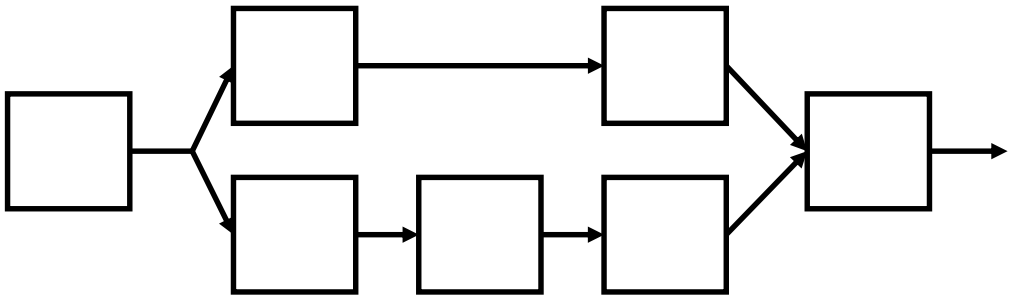
(anylogic)

5.4 Prescriptive Methods

Simulation (examples)

Markov models or *queuing models* are models for analyzing queues where the queue lengths and the waiting time can be determined.

Example: production processes



5.4 Prescriptive Methods

Simulation

Simulation (examples)

Agent-based Modelling (ABM) simulates the actions and interactions of autonomous agents which are assigned with rules that simulate the behavior.

Example: Product launch in the market where the behavior of the customers as well as the competitors are simulated to analyze different scenarios.

5.4 Prescriptive Methods

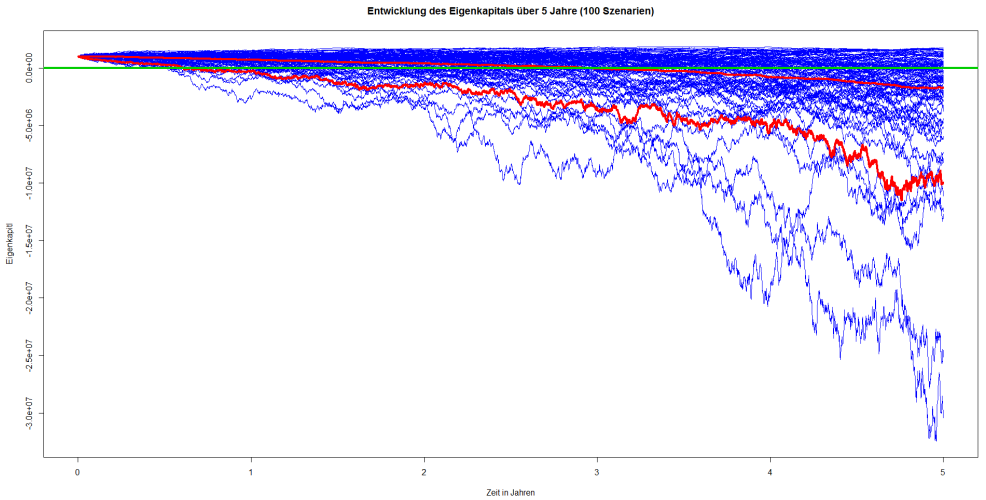
Simulation

Simulation (examples)

Monte Carlo simulation is based on generated random samples which follows (parametric or non-parametric) statistical distributions and interdependencies.

It is typically used to understand the impact of risk and uncertainty in financials of a company e.g. dynamic financial analysis (DFA).

Example: Financial risk models of banks and insurance companies e.g. development of required capital or equity.



5.4 Prescriptive Methods

Simulation

Simulation (examples)

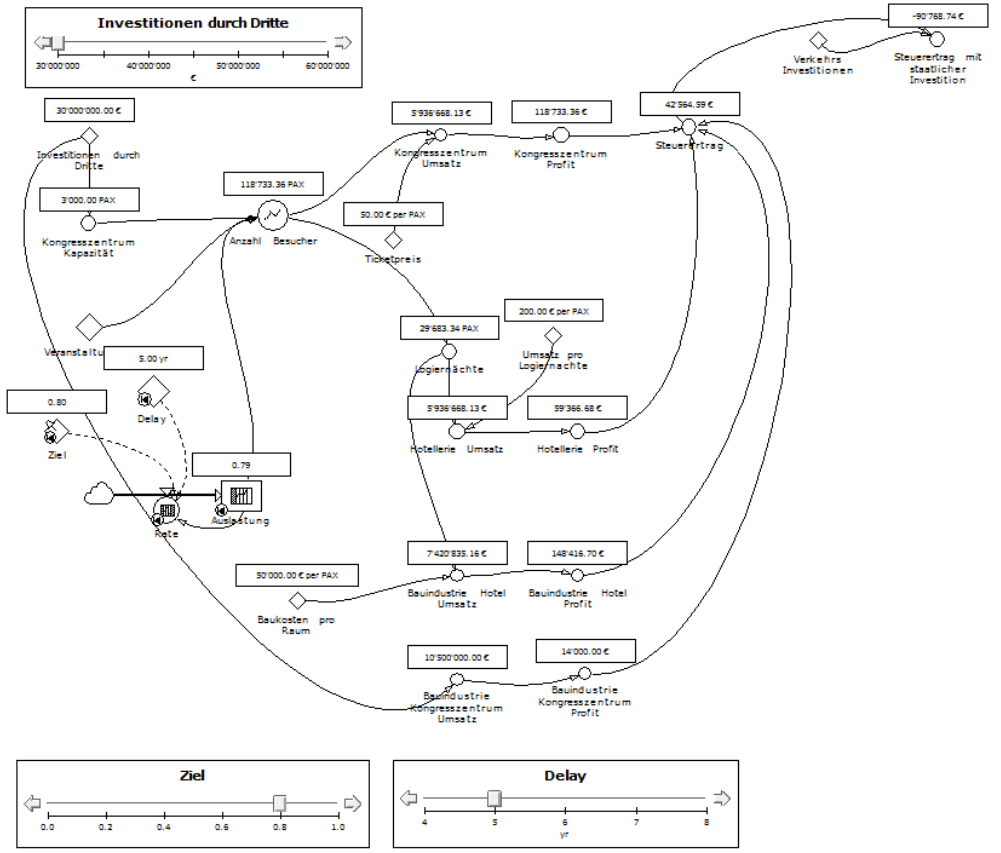
System Dynamics (SD) is a simulation method for the understanding of a complex dynamic system, i.e. the interactions over time. The interactions are not only in a forward oriented direction but can also comprise feedback loops.

5.4 Prescriptive Methods

Simulation

Simulation (examples) (cont'd)

Example: Simulation of a location strategy of a country



5.4 Prescriptive Methods

Simulation

Simulation (examples)

And there are also other simulation methods based on *ordinary / partial differential equations*, *Game Theory*, *stochastic differential equations*, *yield management*, …