# Ausgewählte mathematische Methoden in der Betriebswirtschaft

von

Christian Führer
Frank Hubert
Ralf Gerhards
Jörg Baumgart
Thomas Holey

Schriftleitung: Eva Mroczek

**MANNHEIMER BEITRÄGE ZUR BETRIEBSWIRTSCHAFTSLEHRE**

# Vorwort

## Das Jahr der Mathematik

Zwölf Monate lang stand im Jahr 2008 die Mathematik im Mittelpunkt – als faszinierende Wissenschaft, als ständige Begleiterin in Beruf und Alltag. Ziel des Wissenschaftsjahres 2008 war es, der Öffentlichkeit die Faszination der Mathematik näher zu bringen.

## Mathematik ist facenttenreich

Keine andere Wissenschaft durchdringt und beeinflusst sämtliche Lebens- und Arbeitsbereiche so stark: Vom Automobilbau bis zur Straßenplanung, vom Einkauf im Supermarkt bis zur Architektur, vom Wetterbericht bis zum MP3-Player, vom Bahnverkehr bis zum Internet – alles ist (auch) Mathematik.

## Mathematik ist die Basis

Die Mathematik spielt eine zentrale Rolle in der Wirtschaft und begleitet uns in Alltag und Beruf. Mathematik hilft, Probleme zu analysieren, zu strukturieren und zu lösen. Mit ihren Methoden lassen sich große Teile unserer Lebenswirklichkeit erfassen und beschreiben und viele Phänomene voraussagen.

An der Fakultät für Wirtschaft an der Dualen Hochschule Baden-Württemberg Mannheim haben wir das Mathematik-Jahr zum Anlass genommen, um ausgewählte mathematische Themen und ihre Bedeutung für die Betriebswirtschaftslehre im Einzelbeiträgen darzustellen.

Wir wünschen bei der Lektüre viel Vergnügung.

Rainer Beedgen,
Dekan und Prorektor

# Inhaltsverzeichnis

*Ralf Gerhards*

# Prämienkalkulation in der Lebensversicherung: Wo Betriebswirtschaftslehre, Rechtswissenschaft und Mathematik zusammenfinden

*Von Christian Führer*

## 1 Funktionsweise der Lebensversicherung

Die Lebensversicherungswirtschaft bietet Schutz vor den finanziellen Folgen einschneidender Ereignisse und Entwicklungen im Leben eines Menschen wie Berufs- oder Erwerbsunfähigkeit, Pflegebedürftigkeit, Alter oder Tod. Lebensversicherung definiert sich damit im modernen Sinne des Wortes als privatwirtschaftliche Ergänzung sozialstaatlicher Sicherungssysteme, insbesondere der gesetzlichen Rentenversicherung. Die Idee einer professionellen privatwirtschaftlichen Vorsorge mit Lebensversicherungscharakter ist jedoch schon deutlich älter und kann bis ins 17. Jahrhundert zurückverfolgt werden. Waren die ersten Lebensversicherer noch rein regional agierende Selbsthilfeorganisationen ihrer Versicherungsnehmer, bildeten sich in der ersten Hälfte des 19. Jahrhunderts überregional tätige Lebensversicherer mit zahlreichen Filialbetrieben und großen Außendienstorganisationen, in vielen Fällen die direkten Vorläufer heutiger Lebensversicherungsunternehmen. Im Jahre 2007 waren bei der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) insgesamt 100 Lebensversicherungsunternehmen gemeldet, hinzu kamen einige Lebensversicherer unter Landesaufsicht und zahlreiche ausländische Lebensversicherer, die mit Niederlassungen oder im Rahmen eines grenzüberschreitenden Dienstleistungsgeschäfts auf dem deutschen Markt präsent sind.

Im Unterschied zur gesetzlichen Rentenversicherung orientiert sich die Prämienhöhe in der Lebensversicherung am jeweils vereinbarten Versicherungsschutz, nicht am Bruttoeinkommen des Versicherten. Zentrale Bemessungsgrößen sind die individuelle Art des versicherten Risikos sowie die Höhe des gewünschten Versicherungsschutzes. Die vom Versicherungsnehmer zu entrichtende Prämie wird dabei so kalkuliert, dass der eigene Versicherungsvertrag „im stochastischen Mittel" finanziert ist. In der betrieblichen Praxis ordnet der Lebensversicherer einen neuen Versicherungsnehmer zu diesem Zweck einem Kollektiv gleichartiger Risiken zu, dessen Prämienaufkommen (zuzüglich daraus erwirtschafteter Kapitalerträge) alle in der Zukunft statistisch erwarteten Versicherungsschäden in diesem Kollektiv finanzieren muss. Durch ihren Kapitaldeckungscharakter ist die private Lebensversicherungswirtschaft – im Unterschied zu gesetzlichen Sicherungssystemen – weitgehend unabhängig von der demografi-

schen Entwicklung in der Gesellschaft oder Veränderungen in der Erwerbstätigenstruktur (zum Beispiel vom Verhältnis von Erwerbstätigen zu Nichterwerbstätigen).

Da Lebensversicherungsverträge meist über mehrere Jahrzehnte laufen, müssen die eingezahlten Prämien der Versicherungsnehmer bis zu einem etwaigen Versicherungsfall „zwischengeparkt" werden. Die Versicherungsunternehmen investieren die eingezahlten Prämien hierfür auf dem freien Kapitalmarkt nach strengen gesetzlichen Vorgaben (im Wesentlichen beschrieben in § 54 des Versicherungsaufsichtsgesetzes [VAG] und in der Anlageverordnung [AnlV]), Hauptanlagegrundsätze sind dabei Sicherheit, Rentabilität und Liquidität, was nach § 54 (1) VAG vor allem durch eine hinreichende „Mischung und Streuung" der Kapitalanlagen erreicht werden soll. Im Ergebnis betreibt der Lebensversicherer damit wechselseitige Finanzierungsgeschäfte im Sinne der Theorie der Finanzintermediation (*Kaiser*), fungiert also als Bindeglied zwischen Versichertengemeinschaft und den Kapitalmärkten.

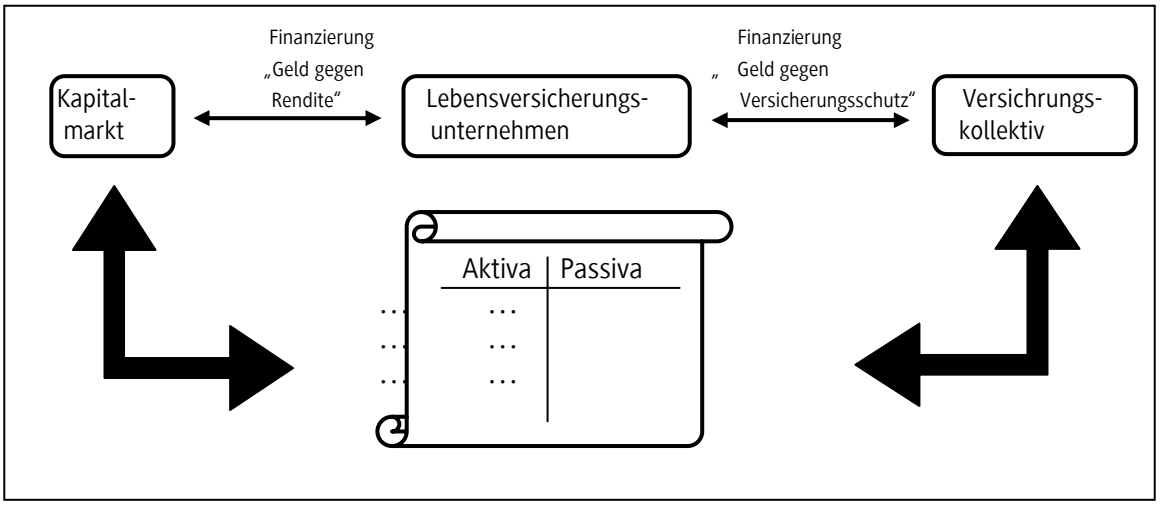



Abb.1: Versicherung als wechselseitiges Finanzierungsgeschäft

## 2 Gesetzliche Vorgaben zur Prämienkalkulation in der Lebensversicherung

Das hauptsächlich im Versicherungsaufsichtsgesetz (VAG) und Versicherungsvertragsgesetz (VVG) formulierte deutsche Versicherungsrecht kann in weiten Teilen als ausgesprochenes Kundenschutzrecht interpretiert werden, das die finanziellen Interessen der Versichertengemeinschaft bewusst vor die finanziellen und anderweitigen Interessen der übrigen Stakeholder am Versicherungsunternehmen (Anteilseigner, Mitarbeiter etc.) stellt. Aufgrund der großen Bedeutung der Lebensversicherung für die private Lebensplanung der Versicherten ist dieser schutztheoretische Ansatz der Versicherungsaufsicht im Bereich der Lebensversicherung besonders deutlich ausgeprägt und findet seinen Niederschlag in zahlreichen Einzelvorgaben an Lebensversicherungsunternehmen, die dadurch in ihrem betriebswirtschaftlichen Gestaltungsspielraum stark eingeschränkt werden.

Die wichtigsten Regelungen speziell für die Prämienkalkulation sind:

- Nach § 11 (1) VAG müssen die Prämien grundsätzlich so kalkuliert sein, dass „*das Versicherungsunternehmen allen seinen Verpflichtungen nachkommen . . . kann*" (Vorsichtsprinzip). Mit Blick auf die Langfristigkeit der Vertragsverhältnisse in der Lebensversicherung verpflichtet das Vorsichtsprinzip die Lebensversicherer dazu, ihre Rechnungsgrundlagen (siehe Abschnitt 3) sehr zurückhaltend zu wählen. Einem allzu aggressiven Wettbewerb (etwa in Form eines „Preiskrieges" zur Gewinnung von Neukunden) unter den Unternehmen wird damit vorgebeugt.
- Infolge des Vorsichtsprinzips erwirtschaften die Lebensversicherer generell hohe Überschüsse; diese müssen an die Versichertengemeinschaft im Rahmen eines relativ komplizierten Überschussbeteiligungsverfahrens zurücktransferiert werden. Details hierzu finden sich beispielsweise in der Verordnung über die Mindestbeitragsrückerstattung (MindZV). Diese rechtliche Vorgabe wird nicht in der Prämienkalkulation direkt angewendet, fordert aber de facto eine laufende a posteriori Überprüfung und Korrektur der Prämienkalkulation.
- Der in § 11 (2) VAG formulierte Gleichbehandlungsgrundsatz besagt schließlich, dass Prämien und Leistungen bei gleichen Voraussetzungen nach gleichen Grundsätzen bemessen sein müssen.

Zusammengenommen erzwingen die genannten gesetzlichen Regelungen in der Lebensversicherungspraxis ein zweistufiges Prämienkalkulationsverfahren, bei dem der Versicherungsnehmer zunächst eine unter Vorsichtsaspekten kalkulierte „zu hohe" Versicherungsprämie entrichtet, nach Ablauf einer festen Zeitperiode (in der Regel ein Jahr) dann aber mit Rückerstattungen im Zuge einer Überschussbeteiligung rechnen kann. Die folgende Darstellung beschäftigt sich ausschließlich mit der Berechnung der zunächst zu entrichtenden Prämie, ein Überblick über die komplizierten Mechanismen der Überschussbeteiligung findet sich etwa in *Führer / Grimmer*.

Die zentralen Rechnungsgrundlagen der Lebensversicherung sind das versicherte Risiko selbst, der zugrunde gelegte Rechnungszins und die vom Versicherer veranschlagten Kosten für die Bereitstellung des Versicherungsschutzes. Das versicherte Risiko wird in der Lebensversicherung je nach Versicherungsprodukt über Sterbewahrscheinlichkeiten, Überlebenswahrscheinlichkeiten oder verschiedene Arten von Invalidisierungswahrschein-lichkeiten abgebildet, die sich in der Regel jeweils am Alter und am Geschlecht der versicherten Person orientieren. Im einfachsten Falle einer Versicherung auf den Tod der versicherten Person werden so genannte Sterbetafeln verwendet, in denen sich die Sterbewahrscheinlichkeiten $q_x$ für Männer bzw. $q_y$ für Frauen finden. Die Variable x bzw. y bezeichnet dabei das Alter einer Person. Werden nur Männer betrachtet, steht $q_{52}$ daher für die Wahrscheinlichkeit eines 52-jährigen Mannes, vor seinem nächsten Geburtstag zu versterben. Die häufig ebenfalls benötigten Überlebenswahrscheinlichkeiten lassen sich aus $q_x$-Werten problemlos herleiten. So beträgt die Wahrscheinlichkeit $p_x$, dass ein x-jähriger Mann seinen nächsten Geburtstag erlebt, gerade

$$p_x = 1 - q_x ,$$

ergibt sich also als Komplementärwahrscheinlichkeit zur Sterbewahrscheinlichkeit $q_x$. Der gleiche x-jährige Mann überlebt zwei Jahre mit einer Wahrscheinlichkeit von (Anwendung des Multiplikationssatzes der elementaren Wahrscheinlichkeitsrechnung, siehe etwa *Kohn*):

$$_2p_x = p_x \cdot p_{x+1} = (1 - q_x) \cdot (1 - q_{x+1})$$

und entsprechend k Jahre mit einer Wahrscheinlichkeit von

$$_kp_x = p_x \cdot p_{x+1} \cdot \ldots \cdot p_{x+k-1} = (1 - q_x) \cdot (1 - q_{x+1}) \cdot \ldots \cdot (1 - q_{x+k-1}) .$$

Mit Hilfe dieser Größen kann dann beispielsweise ermittelt werden, mit welcher Wahrscheinlichkeit ein x-jähriger Mann nach genau k Jahren versterben wird. Hierzu muss er zunächst k Jahre überleben und dann vor Erreichen seines nächsten Geburtstages versterben, was mit einer Wahrscheinlichkeit von

$${}_kp_x \cdot q_{x+k}$$

geschehen wird. Durch diese elementaren wahrscheinlichkeitstheoretischen Überlegungen werden Lebensversicherer in die Lage versetzt, das von einem Versicherungsbestand in k Jahren im stochastischen Mittel benötigte Kapital grob abzuschätzen – eine Grundvoraussetzung jedweder Prämienkalkulation, da alle Versicherungsleistungen letztlich aus Prämienzahlungen und daraus erwirtschafteten Kapitalerträgen finanziert werden müssen. Die Stochastik kann freilich immer nur Erwartungswerte liefern, die tatsächliche Höhe der in k Jahren von einem Lebensversicherer zu erbringenden Versicherungsleistungen kann durchaus höher liegen, was mit Blick auf das Vorsichtsprinzip entsprechende Sicherheitszuschläge bei der Prämienhöhe erforderlich macht.

Da Lebensversicherungsverträge in der Regel über mehrere Jahrzehnte laufen und die Versicherungsleistung erst – wenn überhaupt – zu einem a priori unbekannten Zeitpunkt in der Zukunft fällig wird, investieren die Lebensversicherer nicht benötigte Prämienanteile nach den strengen Vorgaben des § 54 VAG und der Anlageverordnung auf den internationalen Kapitalmärkten. Dieser Verzinsungsprozess muss bei der Prämienkalkulation mit berücksichtigt werden, was im Ergebnis dazu führt, dass künftige Zahlungsströme zwischen Versicherungsnehmern und Versicherungsunternehmen abgezinst („diskontiert") werden müssen. Der hierbei verwendete Zins wird von Seiten des Gesetzgebers nicht explizit vorgeschrieben, die Vorgaben der Deckungsrückstellungsverordnung (DeckRV) legen jedoch eine Verwendung des dort vom Bundesfinanzminister speziell für die Deckungsrückstellung festgelegten Höchstrechnungszinses auch für die Prämienkalkulation nahe. Dieser jährliche Rechnungszins beträgt gegenwärtig (Stand 01.07.2008) 2,25% und ist seit dem Jahre 2000 in zwei Schritten von ehemals 4,0% abgesenkt worden. In der Regel erzielen die Lebensversicherer freilich Kapitalmarkterträge von deutlich über 2,25%. Die dabei erwirtschafteten Zinsüberschüsse fließen ähnlich wie die Risikoüberschüsse (entstanden durch Verwendung „pessimistischer" $q_x$-Werte) im Rahmen der jährlichen Überschussdeklaration fast vollständig an die Versichertengemeinschaft zurück.

Die dritte zentrale Rechnungsgröße sind die zu berücksichtigenden Kosten, die in der Lebensversicherungsmathematik normalerweise in Abschluss-, Inkasso- und Verwaltungskosten aufgeteilt werden (in der Versicherungsmathematik umgangssprachlich auch als „$\alpha$-, $\beta$- und $\gamma$-Kosten" bezeichnet). Auch hier findet das Vorsichtsprinzip des Versicherungsaufsichtsgesetzes Anwendung, weshalb die Kosten in der Regel bewusst zu hoch angesetzt werden, was im Nachhinein zu entsprechenden Kostenüberschüssen zugunsten der Versichertengemeinschaft führt.

# 3 Versicherungsmathematisches Äquivalenzprinzip und Nettoprämienberechnung

Grundlage jedweder Prämienberechnung in der Lebensversicherung ist das versicherungsmathematische Äquivalenzprinzip, nach dem die vom Versicherer gewährten Versicherungsleistungen stets äquivalent zu den vom Versicherungsnehmer gezahlten Prämien sein müssen. Hauptproblem bei der Herstellung dieser Äquivalenz ist die zeitliche Dimension, da der Versicherungsvertrag normalerweise über mehrere Jahrzehnte läuft, womit Verzinsungseffekten eine große Rolle bei der Kapital- und Reservenbildung in der Lebensversicherung zukommt. Die Äquivalenz von Versicherungsleistungen und Prämienzahlungen läuft mathematisch gesprochen auf eine Gleichheit der zugehörigen Barwerte (Gegenwartswerte) hinaus. Da alle Zahlungsströme im Zeitpunkt des Vertragsabschlusses in der Zukunft liegen, müssen zur Berechnung einer Nettoprämie (Prämie ohne Berücksichtigung von Kosten) alle Prämienzahlungen und Versicherungsleistungen mit einem Rechnungszins diskontiert werden. Aus der Gleichheit von Prämienbarwert und Leistungsbarwert lässt sich die Versicherungsprämie dann relativ einfach berechnen.

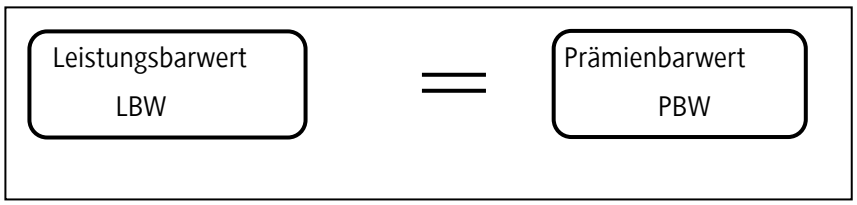



Abb. 2: Versicherungsmathematisches Äquivalenzprinzip

Als Beispiel für eine Nettoprämienberechnung sei an dieser Stelle ein Standardprodukt der Lebensversicherungswirtschaft, die Risikolebensversicherung mit n-jähriger Laufzeit exemplarisch betrachtet. Der Versicherungsnehmer entrichtet bei dieser Versicherungsform während der Vertragslaufzeit zum Beispiel jährlich oder monatlich vorschüssige Prämien (Einmalprämien zu Vertragsbeginn sind auch möglich); verstirbt er während der Vertragslaufzeit, wird die Versicherungsleistung (vereinbarte Versicherungssumme) zum jeweils folgenden Jahresende nachschüssig an die Hinterbliebenen ausgezahlt.

Um den zugehörigen Leistungsbarwert zu ermitteln, wird zunächst die Laufzeit n = 1 betrachtet. In diesem Falle wird die Versicherungssumme VS am Jahresende mit einer Wahrscheinlichkeit $q_x$ fällig, ohne Betrachtung von Verzinsungseffekten muss damit pro Versicherungsnehmer im stochastischen Mittel ein Geldbetrag $VS \cdot q_x$ zur Verfügung stehen, der gleichzeitig einer zu Jahresbeginn einmalig zu entrichtenden Jahresnettoprämie P entspricht:

$$P = VS \cdot q_x .$$

Gerade weil die Versicherungsprämie P bereits zu Jahresbeginn fällig wird, kann dieser Betrag aber genau einmal mit einem jährlichen Rechnungszins i verzinst werden. Dieser Effekt reduziert die tatsächlich zu entrichtende Jahresnettoprämie, nun wird nur noch ein Betrag

$$P = VS \cdot q_x \cdot v$$

fällig, wobei $v = (1 + i)^{-1}$ der zum Rechnungszins i gehörige Diskontierungsfaktor ist. Anders ausgedrückt, muss der Geldbetrag $P = VS \cdot q_x \cdot v$ zu Jahresbeginn vorliegen, damit am Jahresende für die dann im stochastischen Mittel angefallenen Leistungsfälle nach einem Verzinsungszyklus genügend Kapital vorhanden ist.

Für eine allgemeine n-jährige Risikolebensversicherung werden die Leistungsbarwerte für die Einzeljahre jeweils separat berechnet und dann addiert. Für das k-te Jahr beträgt der Leistungsbarwert etwa

$$VS \cdot {}_kp_x \cdot q_{x+k} \cdot v^{k+1} \quad , k = 0, 1, 2, \ldots$$

Dieser Barwert zum Zeitpunkt 0 würde die Versicherungsleistung im Jahre k im stochastischen Mittel finanzieren, das heißt, steht zum Zeitpunkt 0 ein Geldbetrag in dieser Höhe zur Verfügung, wird daraus nach k Jahren mit entsprechender Verzinsung ein Geldbetrag von

$$VS \cdot {}_kp_x \cdot q_{x+k} \cdot v^{k+1} \cdot (1 + i)^{k+1} = VS \cdot {}_kp_x \cdot q_{x+k} \,,$$

also ein Ausdruck der Bauart „VS mal Wahrscheinlichkeit, dass VS überhaupt benötigt wird".

Für den Leistungsbarwert LBW des Gesamtprodukts Risikolebensversicherung mit n-jähriger Laufzeit und konstanter Versicherungssumme VS ergibt sich damit (beachte, dass trivialerweise ${}_0p_x \equiv 1$ gilt):

$$LBW = VS \cdot \sum_{k=0}^{n-1} {}_kp_x \cdot q_{x+k} \cdot v^{k+1},$$

der direkt einer Einmalprämie zum Zeitpunkt 0 entspricht. Oder anders ausgedrückt: Würde der Versicherungsnehmer im Zeitpunkt 0 (Moment des Vertragsabschlusses) eine einmalige Prämienzahlung P = LBW leisten, wäre seine Police für die nächsten n Jahre finanziert. Würde die versicherte Person während dieser n Jahre versterben, stünde die vereinbarte Versicherungssumme VS für die Hinterbliebenen bereit. Die zu zahlende Einmalprämie ist gleich dem Prämienbarwert, da beide zum Zeitpunkt 0 ermittelt werden. Die Gleichung

$$P = VS \cdot \sum_{k=0}^{n-1} {}_kp_x \cdot q_{x+k} \cdot v^{k+1}$$

entspricht damit dem versicherungsmathematischen Äquivalenzprinzip.

Derartige Einmalzahlungen zur Finanzierung einer Lebensversicherung über mehrere Jahrzehnte sind in Deutschland relativ unüblich, was vor allem an der relativ hohen Einmalprämie P liegt. Verbreiteter sind Modelle mit laufender Prämienzahlung, bei denen der Versicherungsnehmer über n Jahre hinweg eine konstante Jahresprämie zahlt (die in einem zweiten Schritt auf Monatsebene herunter gebrochen wird, da die Zahlung in der Praxis meist monatlich erfolgt).

Der hierfür benötigte Prämienbarwert bei laufender Prämienzahlung ergibt sich durch eine ähnliche Überlegung wie beim Leistungsbarwert. Die Nominalwerte aller Prämienzahlungen müssen mit Hilfe eines geeigneten Zinssatzes diskontiert werden; dabei ist zu beachten, dass Prämienzahlungen naturgemäß nur dann erfolgen, wenn der Versicherungsnehmer selbst den Zahlungstermin erlebt. Die einzelnen Prämienzahlungen der (noch unbekannten) Höhe P haben folgende Barwerte:

1. Prämienzahlung: $P \cdot {}_0p_x \cdot v^0 = P$ (Einlösebeitrag bei Vertragsabschluss, keine Diskontierung)
2. Prämienzahlung: $P \cdot {}_1p_x \cdot v^1$
3. Prämienzahlung: $P \cdot {}_2p_x \cdot v^2$

. . .

Insgesamt ergibt sich für eine Vertragslaufzeit von n Jahren bei jährlicher Prämienzahlung damit ein Prämienbarwert

$$PBW = P \cdot \sum_{k=0}^{n-1} {}_kp_x \cdot v^k,$$

alle Prämien werden dabei – wie in der Versicherungswirtschaft üblich – vorschüssig entrichtet, also zu Beginn jedes Versicherungsjahres. Daher tritt der Diskontierungsfaktor v im Prämienbarwert mit einer Potenz k, im Leistungsbarwert jedoch mit einer Potenz (k + 1) auf – Versicherungsleistungen werden nachschüssig erbracht und durchlaufen damit einen zusätzlichen Verzinsungszyklus.

Werden nun Prämien- und Leistungsbarwert gleichgesetzt, ergibt sich:

$$P \cdot \sum_{k=0}^{n-1} {}_kp_x \cdot v^k = VS \cdot \sum_{k=0}^{n-1} {}_kp_x \cdot q_{x+k} \cdot v^{k+1}.$$

Die Summen auf der linken bzw. rechten Seite der Gleichung werden in der Versicherungsmathematik gerne mit $\ddot{a}_{x,n}$ bzw. ${}_nA_x$ abgekürzt, also

$$P \cdot \ddot{a}_{x,n} = VS \cdot {}_nA_x.$$

Die Jahresnettoprämie der n-jährigen Risikolebensversicherung mit laufender jährlicher Prämienzahlung errechnet sich damit zu

$$P = VS \cdot \frac{{}_nA_x}{\ddot{a}_{x,n}}.$$

*Beispiel: Für einen 35-jährigen Mann mit 20-jähriger Vertragslaufzeit und einer Versicherungssumme von 100.000 € ergibt sich bei Verwendung der Sterbetafel DAV 1994 T (in der Branche übliche Sterbetafel der Deutschen Aktuarsvereinigung e.V.) eine Jahresnettoprämie von 418,67 €. Wird der gleiche Versicherungsschutz nur für fünf Jahre benötigt, sinkt die Jahresnettoprämie auf 197,70 €, da der Versicherungsfall nun deutlich unwahrscheinlicher wird. Möchte hingegen ein 65-jähriger Mann eine Risikolebensversicherung für fünf Jahre abschließen, erfordert dies eine Jahresnettoprämie von 3.172,62 €, da nun Lebensjahre mit einer bekanntermaßen hohen Sterblichkeit versichert werden.*

Das vorgestellte Kalkül für die Nettoprämienberechnung bei der n-jährigen Risikolebensversicherung kann problemlos auf andere Versicherungsformen ausgedehnt werden, hier ändert sich dann jeweils nur der Leistungsbarwert. Auch müssen Laufzeit und Prämienzahlungsdauer nicht unbedingt übereinstimmen, in diesen Fällen treten dann in den Summen für den Prämien- und Leistungsbarwert unterschiedliche Parameter (Dauern) n auf. In allen Fällen wird jedoch konsequent das Äquivalenzprinzip der Versicherungsmathematik unter Beachtung stochastischer und finanzmathematischer Gegebenheiten angewendet.

Die Nettoprämie stellt den reinen Versicherungsschutz ohne Beachtung der betrieblichen Gegebenheiten privatwirtschaftlicher Versicherungsunternehmen dar, insbesondere werden die Aufwendungen des Versicherungsunternehmens zur Unterhaltung seiner betrieblichen Infrastrukturen vollständig ausgeblendet. Im Sinne von *Farny* beschränkt sich die Nettoprämie damit auf eine Betrachtung des Risiko- und Kapitalanlagegeschäfts der Versicherer, vernachlässigt jedoch das wichtige Dienstleistungsgeschäft, das unter anderem umfangreiche Beratungs-, Service- und Schadenregulierungstätigkeiten beinhaltet. Folgt man dieser *Farny*'schen Sichtweise, können alle dem Versicherungsunternehmen entstehenden Aufwendungen neben den rein versicherungstechnischen Aufwendungen als Teil der Versicherungsleistungen interpretiert und folglich direkt in den Leistungsbarwert integriert werden. Die Lebensversicherungsmathematik steht damit aber in gewissem Gegensatz zur Rechnungslegung: Dort wird zwischen Aufwendungen für Versicherungsfälle und Aufwendungen für den Versicherungsbetrieb (Abschluss und Verwaltung von Versicherungsverträgen) explizit unterschieden. Für Details sei der Leser auf *Albrecht* oder *Führer / Grimmer* verwiesen.

## 4 Lebensversicherungsmathematik und Versicherungsgeschäft

Die bei laufender Prämienzahlung übliche konstante Prämienhöhe über lange Zeiträume hinweg suggeriert auf den ersten Blick ein zeitlich ebenfalls konstantes versichertes Risiko. Dass dem normalerweise nicht so ist, verdeutlicht ein Blick auf die in der Branche üblicherweise verwendeten Sterbetafeln. In der bereits zitierten Sterbetafel DAV 1994 T finden sich beispielsweise für Männer folgende Sterbewahrscheinlichkeiten:

$q_{20} = 0{,}001476$, $q_{40} = 0{,}002569$, $q_{60} = 0{,}008240$, $q_{80} = 0{,}072101$,

was – wie erwartet – ein mit dem Lebensalter ansteigendes versichertes Risiko für die gewöhnliche Risikolebensversicherung anzeigt. Eine über längere Zeiträume kalkulierte Jahresnettoprämie ist deshalb stets eine Durchschnittsprämie über den betrachteten Zeitraum, das heißt, der Versicherungsnehmer zahlt in jungen Jahren eine an sich zu hohe Prämie (hier *keine* Folge des Vorsichtsprinzips), in höherem Alter dafür eine zu niedrige Prämie. Die in den ersten Vertragsjahren zu viel gezahlten Prämienbestandteile überführt der Lebensversicherer in die so genannte Deckungsrückstellung, die damit zu einem Sammelbecken für Prämienbestandteile wird, die erst in späteren Jahren benötigt werden. Die Deckungsrückstellung zeigt für die drei wichtigsten Produktformen der Lebensversicherungswirtschaft bei laufender Prämienzahlungsweise einen charakteristischen zeitlichen Verlauf (siehe Abb. 3).

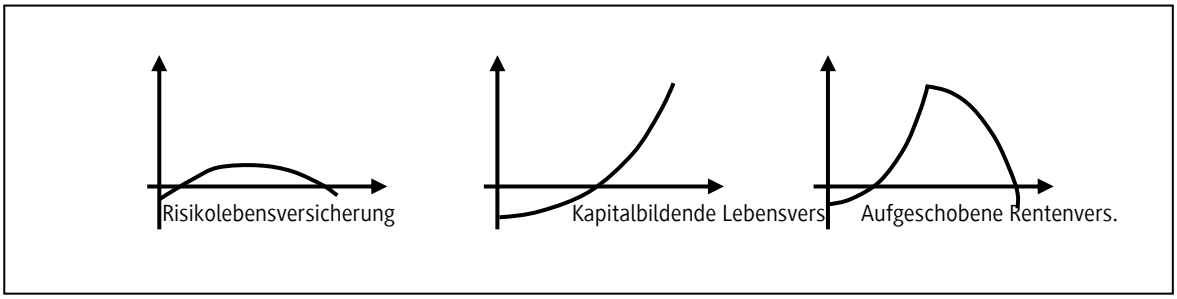



Abb. 3: Zeitlicher Verlauf der Deckungsrückstellung; bei der Risikolebensversicherung erreicht die Deckungsrückstellung nur relativ niedrige Zahlenwerte, da kein Ansparprozess stattfindet

Bei der Risikolebensversicherung werden zunächst Reserven aufgebaut und dann in späteren Jahren aufgezehrt. Das Jahr, in dem sich die Deckungsrückstellung nicht verändert (etwa in der Mitte der Vertragslaufzeit gelegen) ist dasjenige Jahr, in dem die Prämienhöhe genau dem versicherten Risiko (will sagen: der Sterbewahrscheinlichkeit der versicherten Person) entspricht, danach wird die Deckungsrückstellung sukzessive abgebaut und erreicht mit Vertragsende einen Zahlenwert null – eine direkte Folge des Äquivalenzprinzips, das sich ja auf die gesamte Vertragslaufzeit bezieht.

Anders bei der kapitalbildenden Lebensversicherung: Hier wird über die gesamte Vertragslaufzeit hinweg ein Kapitalstock aufgebaut, der mit Ablauf der Versicherung ausgezahlt wird, soweit die versicherte Person nicht schon vorher verstirbt. Bei der aufgeschobenen Rentenversicherung ergibt sich hingegen ein charakteristischer Sägezahnverlauf. Zunächst wird ein Kapitalstock aufgebaut, während der Rentenbezugsphase dann langsam wieder aufgezehrt. Bei dieser Versicherungsform erreicht die Deckungsrückstellung in genau dem Jahr den Zahlenwert null, in dem die versicherte Person im stochastischen Mittel versterben würde – ein Zahlenwert, der in den vergangenen Jahrzehnten stetig angestiegen ist. Die zuvor verstorbenen versicherten Personen finanzieren indirekt die Rentenzahlungen an diejenigen Versicherten, die „länger als erwartet" leben.

Eine einfache Modellrechnung soll die Funktionsweise der Deckungsrückstellung verdeutlichen. Schließt ein 40-jähriger Mann eine zweijährige Risikolebensversicherung mit Versicherungssumme VS = 100.000 € ab, entspricht dies nach dem obigen Kalkül bei einem Rechnungszins von 2,25 % (das heißt, v = 0,97799511) einem Leistungsbarwert

$$LBW = VS \cdot {}_0p_x \cdot q_x \cdot v + VS \cdot {}_1p_x \cdot q_{x+1} \cdot v^2 = VS \cdot (q_{40} \cdot v + (1 - q_{40}) \cdot q_{41} \cdot v^2)$$

$$= 520{,}5660438\ldots$$

Dabei wurde die schon erwähnte Sterbetafel DAV 1994 T verwendet. Für den Prämienbarwert bei laufender Prämienzahlung gilt hingegen (${}_1p_x = p_x = p_{40} = (1 - q_{40})$):

$$PBW = P \cdot {}_0p_x \cdot v^0 + P \cdot {}_1p_x \cdot v = P \cdot (1 + (1 - q_{40}) \cdot v)$$

$$= P \cdot 1{,}9754826\ldots ,$$

die Jahresnettoprämie P ist dabei noch unbekannt, sie errechnet sich durch Gleichsetzen der beiden Barwerte (Äquivalenzprinzip) und anschließender Division zu

$$P = 520{,}5660438 : 1{,}9754826 = 263{,}51\ € .$$

Diese Jahresnettoprämie ist zweimal, jeweils zu Beginn des ersten und des zweiten Versicherungsjahres (Zeitpunkte t = 0 und t = 1), zu entrichten und entspricht einer Durchschnittsprämie über die zwei Jahre.

Tatsächlich wird im ersten Jahr im stochastischen Mittel ein Barwert der Höhe

$$VS \cdot q_{40} \cdot v = 251{,}25\ €$$

für die Darstellung des gewünschten Versicherungsschutzes benötigt. Der Differenzbetrag 263,51 – 251,25 = 12,26 € wird entsprechend in die Deckungsrückstellung überführt und einmal mit 2,25% verzinst: 12,26 · 1,0225 = 12,54 €. Nach Ablauf eines Jahres wird nun für die Versicherungsleistung im zweiten Jahr ein Betrag

$VS \cdot q_{41} \cdot v = 276{,}07$ €

benötigt. Tatsächlich zahlt der Versicherungsnehmer aber lediglich wieder P = 263,51 €. Zusammen mit der Deckungsrückstellung gilt nun aber

263,51 + 12,54 = 276,05 €,

sodass der geforderte Leistungsbarwert für das zweite Versicherungsjahr (von einer kleinen Rundungsabweichung einmal abgesehen) in der Tat zur Verfügung steht. Nicht umsonst hieß die Deckungsrückstellung früher „Prämienreserve", speziell in der Risikolebensversicherung mit ihrem zeitlich schwankenden Risikoverlauf ist ihre diesbezügliche Funktion offenkundig.

Die Deckungsrückstellung findet sich als Verbindlichkeit auf der Passivseite der Bilanz eines Lebensversicherungsunternehmens und stellt dort den mit Abstand größten Einzelposten dar. So entfielen im Jahre 2006 immerhin 83,5% der Gesamtbilanzsumme aller bei der Bundesaufsichtsbehörde BaFin gemeldeten Lebensversicherer auf die Deckungsrückstellung, nur 1,5% hingegen auf das Eigenkapital. Die übrigen 15% entfielen im Wesentlichen auf Passivposten mit Bedeutung für die Überschussbeteiligung, nämlich die Rückstellung für die Beitragsrückerstattung (RfB) und verzinslich angesammelte Überschussanteile. Letztlich wird eine versicherungsmathematische Standardvorgehensweise damit zum zentralen Bestimmungsparameter der Rechnungslegung von Lebensversicherungsunternehmen.

## 5 Ausblick

Mit der Deregulierung der Versicherungsmärkte im Jahre 1994 (Umsetzung einer EU-Richtlinie in deutsches Recht) hat sich der Lebensversicherungsmarkt in Deutschland grundlegend verändert. Die Lebensversicherer genießen seither große Freiheiten bei der Produktgestaltung, sehen sich damit aber auch zunehmend der Notwendigkeit gegenüber, diese Freiheiten betriebswirtschaftlich sinnvoll auszufüllen.

Zu den markantesten Produkttrends in der Lebensversicherung gehörten in der jüngsten Vergangenheit zweifelsohne die Preferred Lives-Produkte, bei denen neben dem Alter und Geschlecht der versicherten Person auch noch andere Risikofaktoren wie das Rauch- und Alkoholkonsumverhalten, Blutwerte oder das Betreiben von Risikosportarten mit in die Kalkulation einfließen. Solche Überlegungen erfordern freilich einen anderen Umgang mit Sterbetafeln, insbesondere muss die Lebensversicherungsmathematik Verfahren entwickeln, die eine Abschätzung der Sterbewahrscheinlichkeit bei Kombination mehrerer solcher Risikofaktoren sicher ermöglichen.

Eine weitere Herausforderung an die Lebensversicherungsmathematik stellen aufsichtsrechtliche Überlegungen zu einer besseren Abstimmung von versicherungstechnischen Verbindlichkeiten („Liabilities") und den im Versicherungsunternehmen zu deren Bedeckung vorhandenen Kapitalanlagen („Assets") im Rahmen eines dezidierten Asset-Liability-Managements dar (kurz „ALM", siehe etwa *Jost*). Hier geht es vor allem darum, das Versicherungsunternehmen langfristig in die Lage zu versetzen, seine versicherungstechnischen Verbindlichkeiten jederzeit erfüllen zu können („Solvabilität"), was einem Ausgleich der klassischen finanzwirtschaftlichen Ziele Sicherheit, Rendite und Liquidität gleichkommt. Auf internationaler Ebene werden im Rahmen des Projekts Solvency II gegenwärtig erstmals verbindliche Richtlinien künftiger ALM-Ansätze ausgearbeitet, siehe etwa *Gündl / Perlet* oder *Nguyen*.

Langfristig werden die betriebswirtschaftlichen und rechtlichen Anforderungen an Mathematiker in der Lebensversicherung daher weiter zunehmen, wird es zu einer noch innigeren Verbindung von Betriebswirtschaft, Rechtswissenschaft und Mathematik im Lebensversicherungsunternehmen kommen, was die Lebensversicherungsmathematik auch künftig zu einem spannenden und einträglichen Berufsfeld machen wird.

## Literatur

Albrecht, P. (2007): Grundprinzipien der Finanz- und Versicherungsmathematik, Schäffer-Poeschel Verlag, Stuttgart, 2007

Farny, D. (2006): Versicherungsbetriebslehre, 4. Auflage, Verlag Versicherungswirtschaft, Karlsruhe, 2006

Führer, C.; Grimmer, A. (2006): Einführung in die Lebensversicherungsmathematik, Verlag Versicherungswirtschaft, Karlsruhe, 2006

Gündl, H.; Perlet, H. (2005): Solvency II & Risikomanagement: Umbruch in der Versicherungswirtschaft, Gabler Verlag, Wiesbaden, 2005

Jost, C. (1995): Asset-Liability-Management bei Versicherungen, Schriftenreihe „Versicherung und Risiko" des Instituts für betriebswirtschaftliche Risikoforschung und Versicherungswirtschaft der Ludwig-Maximilians-Universität München, Gabler Verlag, Wiesbaden, 1995

Kaiser, D. (2006): Finanzintermediation durch Banken und Versicherungen – Die theoretischen Grundlagen der Bankassurance, Gabler Verlag, Wiesbaden, 2006

Kohn, W. (2006): Statistik, Springer Verlag, Berlin und Heidelberg, 2005

Nguyen, T. (2008): Handbuch der wert- und risikoorientierten Steuerung von Versicherungsunter-nehmen, Verlag Versicherungswirtschaft, Karlsruhe, 2008

# Cournotscher Punkt oder der „böse" Monopolist

*von Frank Hubert*

## 1 Marktwirtschaft und Wettbewerb

Im Wettstreit der Wirtschaftssysteme hat sich die Marktwirtschaft gegenüber der Zentralverwaltungswirtschaft als überlegen erwiesen. Fast alle wichtigen Volkswirtschaften haben inzwischen marktwirtschaftlich geprägte Wirtschaftsordnungen. Allerdings weisen diese zahlreiche Modifikationen zum System einer reinen Marktwirtschaft auf. Dies gilt auch für die Soziale Marktwirtschaft, die sich in Deutschland nach dem Zweiten Weltkrieg durchgesetzt hat. Allen marktwirtschaftlichen Wirtschaftsordnungen ist allerdings gemein, dass sie die Bedeutung des Wettbewerbs betonen.

Gleichwohl kann man in der Praxis immer wieder feststellen, dass in vielen Fällen kein Wettbewerb stattfindet. Eine Zugfahrt von Mannheim nach Frankfurt ist nur mit der Deutschen Bahn AG möglich. Der Wechsel des Gasversorgers scheitert in vielen Fällen, weil keine oder nur wenige alternative Anbieter ihre Dienste anbieten. Bis Ende 2007 gab es - mit vernachlässigbaren Ausnahmen - ein Briefmonopol im Postdienst. Ähnliches galt für das Telefonieren im Festnetz bis zur Liberalisierung Anfang 1998. Auch wenn eine Reihe von Monopolen inzwischen aufgebrochen wurde, bleibt es auf vielen Märkten zunächst bei einer monopolähnlichen Stellung des etablierten Anbieters, da sich nur sehr langsam potentielle Konkurrenten finden, die es wagen, in den Wettbewerb einzusteigen.

Der folgende Beitrag beschäftigt sich mit der Frage, weshalb Monopole in einer marktwirtschaftlichen Ordnung in aller Regel ein Problem darstellen. Dazu werden zunächst die grundlegenden Gewinn-, Erlös- und Kostenbegriffe erläutert. Danach werden die Marktergebnisse auf einem Monopolmarkt analytisch und graphisch ermittelt und dann den Marktergebnissen bei vollkommener Konkurrenz, einem homogenen Polypol, gegenübergestellt. Dieser Vergleich wird anschließend um Märkte mit wenigen Anbietern, den Oligopolen, erweitert. Die Resultate zeigen die Notwendigkeit einer Wettbewerbspolitik zur Aufrechterhaltung der marktwirtschaftlichen Ordnung.

# 2 Gewinnmaximierung als Unternehmensziel

## 2.1 Gewinn, Erlös und Grenzerlös

Für die Analyse der verschiedenen Marktformen muss zunächst das Ziel der Anbieter definiert werden. In einer Marktwirtschaft ist davon auszugehen, dass die Produzenten die Gewinnmaximierung anstreben. Viele andere Ziele dienen letztlich dazu, einen möglichst hohen Gewinn zu erreichen (vgl. Bartmann, H./Busch, A. A./Schwaab, J. A., 1999, S. 35f.). Dies gilt sowohl für Marktstellungsziele (z. B. hoher Marktanteil) als auch für finanzielle Ziele (z. B. Kreditwürdigkeit) und soziale Ziele (z. B. Zufriedenheit der Mitarbeiter).

Der Gewinn G ist definiert als Differenz aus den Erlösen E und den Kosten K, die beide von der angebotenen Menge x abhängen:

(1) $G(x) = E(x) - K(x) \quad \max.!$

Als notwendige Bedingung für das Gewinnmaximum erhält man somit:

(2) $\frac{dG}{dx} = 0 \quad \Rightarrow \quad \frac{dE}{dx} = \frac{dK}{dx} \quad \text{bzw.} \quad GE = GK\,.$

Der Erlös, der durch eine zusätzliche verkaufte Einheit erzielt wird, entspricht im Gewinnmaximum genau den Kosten, die für eine zusätzliche Einheit des jeweiligen Produkts entstehen. Der Grenzerlös GE ist gleich den Grenzkosten GK.

Für eine ausführlichere Analyse müssen Umsatz und Kosten genauer betrachtet werden. Der Umsatz ist das Produkt aus Menge x und Preis p. Liegt vollkommene Konkurrenz vor, so ist der Preis für den Anbieter vorgegeben. Da dieser nur einer von sehr vielen Anbietern auf einem Markt ist, hat er keinen direkten Einfluss auf den Preis. Der Preis ist für ihn ein Datum. Der Anbieter kann nur seine Menge anpassen.

Völlig anders ist dagegen die Situation des Monopolisten. Da er der einzige Anbieter ist, kann er den Preis bestimmen. Bei einer normal verlaufenden, fallenden Nachfragefunktion, ist der Absatz einer großen Menge nur bei einem niedrigen Preis möglich, während beim Absatz einer kleinen Menge ein sehr hoher Preis verlangt werden kann. Der Preis ist somit abhängig von der Menge, die der Alleinanbieter absetzen will (vgl. Petersen, T., 2008, S. 67). Daher unterscheidet sich die formale Darstellung von Erlös und Grenzerlös des Monopolisten (3a) von denen eines Anbieters im homogenen Polypol (3b):

(3a) $E = p(x) \cdot x \quad \text{und} \quad GE = \frac{dE}{dx} = p + \frac{dp}{dx} \cdot x$

und

(3b) $E = p \cdot x \qquad \text{und} \quad GE = \frac{dE}{dx} = p$

## 2.2 Kosten, Grenzkostipen und Stückkosten

Für die Darstellung der verschiedenen Kostenfunktionen müssen einige Annahmen getroffen werden. Generell kann davon ausgegangen werden, dass bei der Erstellung von Waren und Dienstleistungen sowohl fixe Kosten $K_F$ als auch variable Kosten $K_V$ anfallen. Fixe Kosten, wie z. B. die Miete für die Produktions- und Lagerhallen, sind unabhängig von der produzierten Menge, während die variablen Kosten, wie z. B. die Materialkosten, mit wachsender Menge steigen:

$$(4) \qquad K = K_F + K_V(x) \quad \text{mit} \quad \frac{dK_F}{dx} = 0 \quad \text{und} \quad \frac{dK_V}{dx} > 0 \,.$$

Der genaue Verlauf von Produktionsfunktionen und den daraus hergeleiteten Kostenfunktionen ist in den Wirtschaftswissenschaften umstritten (vgl. Frantzke, A., 2004, S. 142ff., Mankiw, N. G., 2004, S. 301ff. und Varian, H. R., 2007, S. 457ff.). In vielen Unternehmen dürfte die Kostenfunktion den in Abb. 1 dargestellten Verlauf haben.

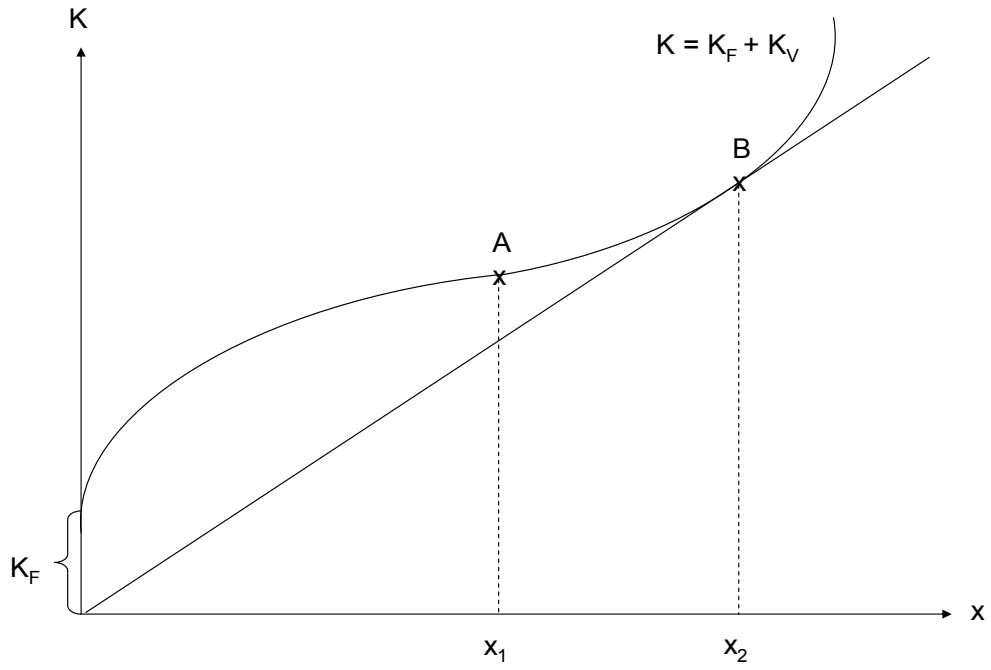



Abb. 1: Kostenfunktion

Dieser Verlauf kann aus einer ertragsgesetzlichen Produktionsfunktion hergeleitet werden. Er bedeutet, dass die Kosten mit steigender Menge zunächst aufgrund zunehmender Skalenerträge nur langsam steigen. Gründe dafür können Optimierungen in der Produktion bei steigender Ausbringungsmenge oder Mengenrabatte der Zulieferer bei größeren Einkaufsmengen sein. Ab einem bestimmten Punkt A, der mathematisch betrachtet den Wendepunkt der Kostenfunktion darstellt, steigen die Kosten deutlich an. Der überproportionale Anstieg kann z. B. dadurch verursacht sein, dass sich zumindest kurzfristig ein Teil der Produktionsfaktoren nicht beliebig vermehren lässt. Es ist nur eine partielle Faktorvariation möglich (vgl. Herdzina, K., 2005, S. 120ff.). Die Unternehmen können bspw. ihren Produktionsstandort aus Platzmangel

nicht weiter ausbauen und daher keine zusätzlichen Maschinen aufstellen. Die Mehrproduktion ist dann nur mit Nacht- und Wochenendschichten der Mitarbeiter möglich. Wegen der Schichtzuschläge steigen die Kosten nun aber deutlich an. Eine andere Erklärung für stark steigende Kosten wäre die in vielen Großbetrieben zu beobachtende Komplizierung der Abläufe durch einen stark steigenden Einfluss der Verwaltung.

Eine derartige Kostenfunktion hat Konsequenzen für die Verläufe der Grenzkostenfunktion GK, der Durchschnittskostenfunktion (synonym: Stückkostenfunktion) k und der Funktion der variablen Durchschnittskosten (synonym: variable Stückkostenfunktion) $k_V$, die wie folgt definiert sind:

(5a) $$GK = \frac{dK}{dx},$$

(5b) $$k = \frac{K}{x} \quad \text{und}$$

(5c) $$k_V = \frac{K_V}{x}.$$

Die Grenzkostenfunktion fällt zunächst, da mit jeder zusätzlichen produzierten Einheit die Zunahme der Kosten geringer wird. Im Wendepunkt A der Kostenfunktion hat die Grenzkostenfunktion ihr Minimum. Notwendige Bedingung hierfür ist, dass die erste Ableitung der Grenzkostenfunktion den Wert Null annimmt:

(6) $$\frac{dGK}{dx} = \frac{d^2K}{dx^2} = 0.$$

Die erste Ableitung der Grenzkostenfunktion stellt aber nichts anderes dar als die zweite Ableitung der Kostenfunktion, die zur Bestimmung des Wendepunkts A benötigt wird. Werden ab der Menge $x_1$ zusätzliche Einheiten produziert, so steigen die Kosten für jede zusätzliche Einheit immer weiter an.

In Abb. 2 ist neben der Grenzkostenfunktion auch die Stückkostenfunktion dargestellt. Das Minimum dieser Funktion erhält man graphisch, indem man in Abb. 1 eine Tangente aus dem Ursprung an die Kostenfunktion legt. Alle Punkte auf dieser Geraden stehen für das gleiche Verhältnis von K zu x und damit für gleiche Durchschnittskosten. Punkte oberhalb des Geraden bedeuten höhere Stückkosten, Punkte unterhalb der Geraden dagegen geringere Stückkosten.

Im Punkt B ist damit das Verhältnis von Kosten K und Menge x kleiner als auf jedem anderen Punkt der Kostenfunktion. Die Durchschnittskosten haben daher an dieser Stelle ihr Minimum. Man bezeichnet diesen Punkt auch als Betriebsoptimum. Abb. 2 zeigt, dass die Stückkosten in ihrem Minimum von der Grenzkostenkurve geschnitten werden. Produziert man ausgehend von der Menge $x_2$ eine Einheit mehr, so liegen die Kosten für diese zusätzliche Einheit über den bisherigen Stückkosten. Damit steigen aber auch der neue Durchschnitt und somit die Durchschnittskostenkurve ab dieser Menge leicht an.

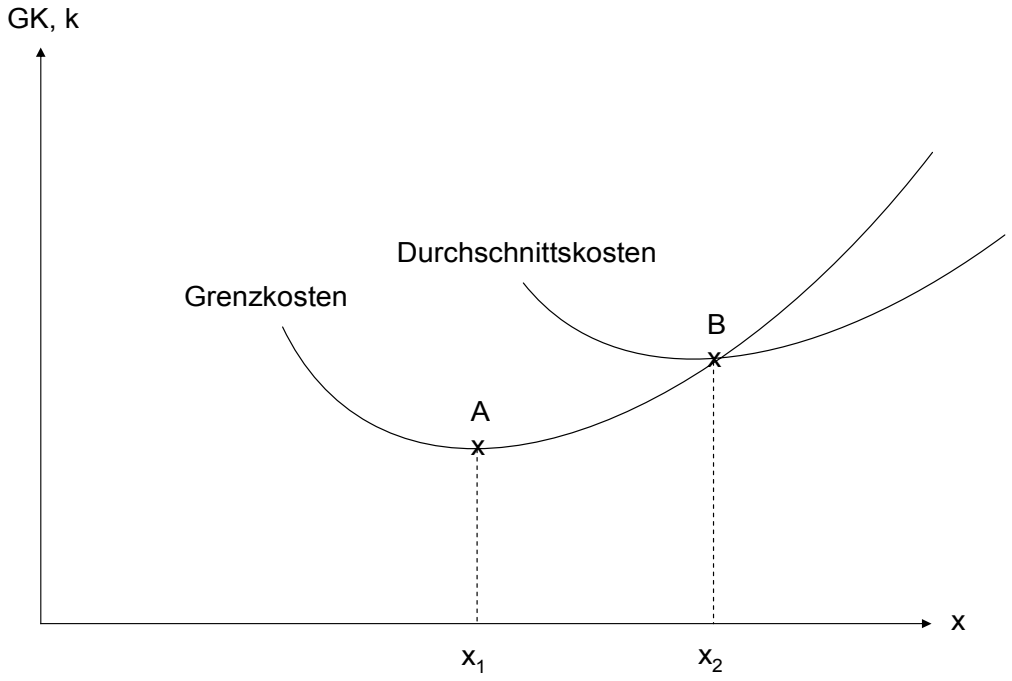



Abb. 2: Grenzkosten- und Durchschnittskostenfunktion

Die Funktion der variablen Stückkosten, die in den Abbildungen nicht berücksichtigt ist, hat ebenfalls einen U-förmigen Verlauf. Sie liegt immer unterhalb der Stückkostenfunktion, da nur ein Teil der Kosten berücksichtigt wird. Im Minimum wird sie ebenfalls von der Grenzkostenkurve geschnitten. Diesen Punkt bezeichnet man als Betriebsminimum. Er spielt bei kurzfristigen Überlegungen eine Rolle, da an diesem Punkt zumindest die variablen Durchschnittskosten gedeckt werden.

# 3 Preisbildung im Monopol

## 3.1 Mathematische und grafische Darstellung

Für den Monopolisten gilt es nun nach den bisherigen Überlegungen die Angebotsmenge zu bestimmen, bei der die Grenzkosten dem Grenzerlös entsprechen. Da der Monopolist zusätzliche Güter nur verkaufen kann, wenn er den Preis senkt, ist der Grenzerlös nicht grundsätzlich positiv. Formal lässt sich dieser Zusammenhang mittels der Preiselastizität der Nachfrage $\eta$ erklären (vgl. Petersen, T., 2008, S. 67), die wie folgt definiert ist:

$$\eta = \frac{dx}{x} : \frac{dp}{p} = \frac{dx}{dp} \cdot \frac{p}{x}. \tag{7}$$

Diese Elastizität gibt an, um wie viel Prozent sich die nachgefragte Menge ändert, wenn der Preis des Gutes um ein Prozent erhöht bzw. gesenkt wird. Bei einer normal verlaufenden Nachfragefunktion liegt der Wertebereich von $\eta$ zwischen 0 und $-\infty$, wobei der Wert 0 nicht erreicht werden kann. So führt bspw. bei einem Wert von -2 eine Preiserhöhung um ein Prozent zu einem Rückgang der nachgefragten Menge von zwei Prozent.

Mittels einiger weniger Umformungen kann ein Zusammenhang zwischen dem Grenzerlös und der Preiselastizität der Nachfrage hergestellt werden. Diese Verknüpfung wurde in der Volkswirtschaftslehre erstmals von Luigi Amoroso (1886-1965) und Joan Robinson (1903-1983) gezogen und ist daher auch unter dem Namen Amoroso-Robinson-Relation bekannt:

$$(8) \qquad GE = \frac{dE}{dx} = p + \frac{dp}{dx} \cdot x \cdot \frac{p}{p} = p + \frac{dp}{dx} \cdot \frac{x}{p} \cdot p = p + p \cdot \frac{1}{\eta} = p \cdot \left(1 + \frac{1}{\eta}\right).$$

Der Grenzerlös hat nur in dem Bereich positive Werte, in dem die Preiselastizität der Nachfrage kleiner als -1 ist. Man spricht in diesem Fall von einer preiselastischen Nachfrage. Ist die Nachfrage dagegen unelastisch, so liegt $\eta$ zwischen 0 und -1. Für den Grenzerlös erhält man dann negative Werte. Da im Gewinnmaximum Grenzkosten und Grenzerlöse übereinstimmen müssen, kommt wegen der positiven Grenzkosten nur der elastische Bereich als Lösung für das Gewinnmaximierungsproblem in Frage.

Für den Monopolisten ist die gesamtwirtschaftliche Nachfragefunktion nichts anderes als seine persönliche Preis-Absatzfunktion (vgl. Bartmann, H./ Busch, A. A./Schwaab, J. A., 1999, S. 119). Geht man vereinfachend von einem linearen Zusammenhang zwischen dem Preis und der nachgefragten Menge aus, so kann man dies über Gleichung (9) beschreiben:

$$(9) \qquad p = a - b \cdot x \quad \text{mit } a, b > 0.$$

Unter Verwendung von (3a) erhält man für den Erlös bzw. den Grenzerlös:

$$(10a) \qquad E = (a - b \cdot x) \cdot x = a \cdot x - b \cdot x^2$$

und

$$(10b) \qquad GE = \frac{dE}{dx} = a - 2 \cdot b \cdot x.$$

Der Vergleich der Formeln (9) und (10b) zeigt, dass sowohl die Nachfragefunktion als auch die Grenzerlösfunktion für x = 0 im Punkt a beginnen. Die Steigung ist in beiden Fällen negativ, wobei die Steigung der Grenzerlösfunktion betragsmäßig doppelt so groß ist wie die Steigung der Nachfragefunktion.

Abb. 3 verdeutlicht, welche Konsequenzen dies für die Bestimmung des Gewinnmaximums hat. Der Schnittpunkt von Grenzkosten und Grenzerlös liegt bei der relativ geringen Monopolmenge $x_M$. Würde der Monopolist die Menge erhöhen, so müsste er Preiszugeständnisse machen. Der Grenzerlös wäre zwar zunächst noch positiv, aber kleiner als die Grenzkosten. Dies hätte eine Verringerung des Gewinns zur Folge.

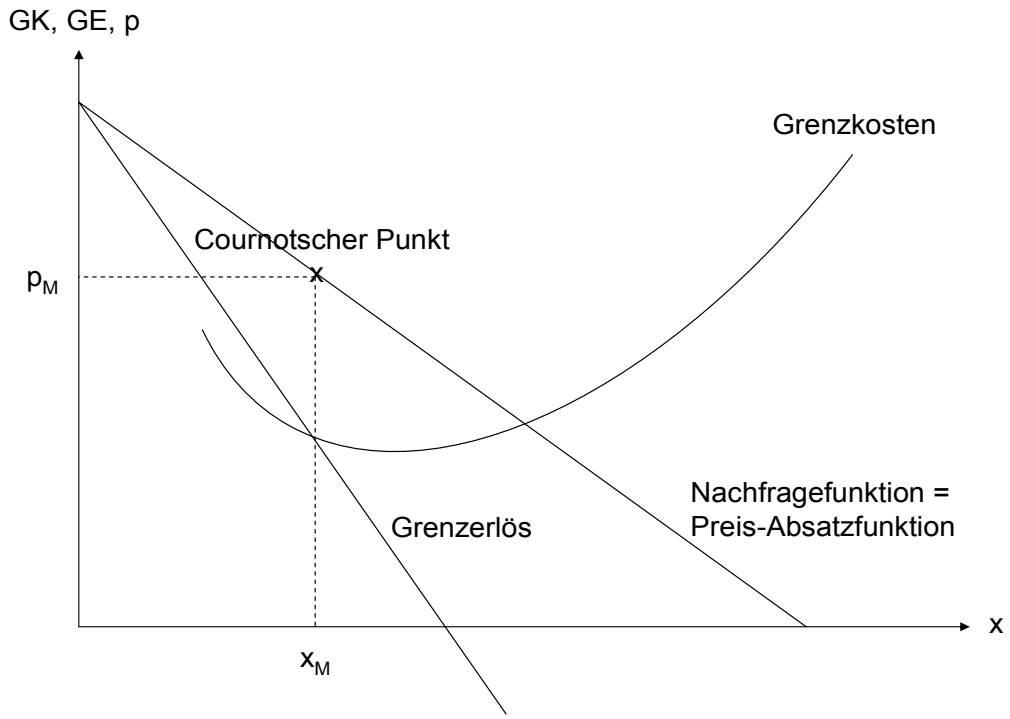



Abb. 3: Bestimmung des Cournotschen Punktes

Die Monopolmenge wird zu einem relativ hohen Monopolpreis $p_M$ angeboten. Diesen Preis erhält man, indem man vom Schnittpunkt der Grenzkosten- und der Grenzerlösfunktion senkrecht nach oben auf die Nachfragefunktion N stößt. Dieser Punkt, der in Erinnerung an den französischen Nationalökonom Antoine-Augustin Cournot (1801-1877) auch als Cournotscher Punkt bezeichnet wird, gibt an, zu welchem Preis die Nachfrager die angebotene Menge abnehmen. Er steht somit für das Marktergebnis, das sich auf einem Markt mit einem gewinnmaximierenden Monopolisten einstellt.

**Leben und Werk Antoine-Augustin Cournots:**
(vgl. Moore, H. L./Waffenschmidt, W. G., 1965 und Zimmermann, L. J. R./Marcon, H., 1989)

Der französische Mathematiker und Wirtschaftstheoretiker Antoine-Augustin Cournot wurde am 28. August 1801 in Gray im Département Haute-Saône geboren. Cournot studierte Mathematik und Naturwissenschaften in Besancon und Paris. Zudem beschäftigte er sich intensiv mit philosophischen Fragestellungen. Auf Vermittlung des bekannten Mathematikers Siméon Denis Poisson wurde er 1834 Professor für Mathematik in Lyon. Weitere Stationen seines Wirkens waren Grenoble, Paris und Dijon. Cournot hatte maßgeblichen Anteil an der Mathematisierung der Wirtschaftswissenschaften. Insbesondere in der Mikroökonomie sind viele seiner Ansätze und Ideen nach wie vor ein Muss für jeden Studierenden, der sich mit volkswirtschaftlichen Fragestellungen beschäftigt. Sein Name ist sowohl mit dem Cournotschen Punkt im Monopolfall als auch mit dem Cournot-Modell für das Duopol verbunden.

In seinem bedeutendsten Werk „Recherches sur les principes mathématiques de la théorie des richesses" (deutsch: „Untersuchungen über die mathematischen Grundlagen der Theorie des Reichtums"), das 1838 erschien, untersucht er die Auswirkungen verschiedener Marktformen auf die Marktergebnisse. Diese werden sowohl analytisch als auch graphisch umfassend dargestellt. Die deutsche Fassung seiner Abhandlung liegt, übersetzt von Walter Gustav Waffenschmidt, seit 1924 vor. Cournot starb am 31. März 1877 in Paris.

## 3.2 Variationen der Monopolpreisbildung

Ob der Cournotsche Punkt in der Praxis immer realisiert wird, hängt allerdings noch von einigen anderen Faktoren ab. Zunächst ist dieser Punkt nur dann relevant, wenn der Monopolpreis mindestens die Stückkosten deckt (vgl. Petersen, T., 2008, S. 67). Ist dies nicht der Fall, erleidet der Monopolist einen Verlust und verzichtet daher völlig auf ein Angebot. Der Stückgewinn g ist in diesem Fall negativ:

(11) $g = p - k$, für $k > p$: Verlust

Muss der Monopolist mit potentiellen Konkurrenten rechnen, kann es ebenfalls sein, dass die Cournot-Lösung zunächst nicht zustande kommt. In diesem Fall versucht der Anbieter vorübergehend den Markt mit niedrigeren Preisen und höheren Mengen zu überschwemmen. Dadurch kann er die potentiellen Anbieter abschrecken und ihren Markteintritt möglicherweise verhindern (vgl. Herdzina, K., 2005, S. 172). Ist dies gelungen, kann der Alleinanbieter anschließend wieder den Cournotschen Punkt realisieren.

Schließlich ist es auch denkbar, dass der Monopolist unterschiedliche Preise von verschiedenen Nachfragergruppen verlangt. In Abb. 3 erkennt man, dass manche Nachfrager auch bereit wären, einen höheren Preis zu zahlen. Gelingt es nun dem Monopolisten durch eine Preisdifferenzierung diese höhere Zahlungsbereitschaft auszunutzen, so kann er seinen Gewinn weiter steigern (vgl. Frantzke, A., 2004, S. 243). Will man z. B. mit dem Zug von Frankfurt nach Mannheim fahren, so geht dies nur mit dem Monopolisten Deutsche Bahn AG. Gleichwohl verlangt die Bahn für diese Strecke nicht von allen Reisenden den gleichen Preis. Es gibt eine Preisdifferenzierung, die davon abhängt, welchen Zug man nutzt (Regionalexpress oder ICE) und ob man erste oder zweite Klasse fährt.

**Ein Zahlenbeispiel zur Monopolpreisbildung:**

Ein Monopolist sieht sich folgender Preis-Absatz-Funktion gegenüber:
$p = 50 - 0{,}5x$ .
Die Kostenfunktion des Monopolisten lautet:

$$K = 10x - 2x^2 + \frac{x^3}{3} + \frac{100}{3}.$$

Somit erhält man für die Grenzkosten- und die Stückkostenfunktion:

$GK = 10 - 4x + x^2$ und

$$k = 10 - 2x + \frac{x^2}{3} + \frac{100}{3x}.$$

Erlös und Grenzerlös lassen sich unter Verwendung der Preis-Absatz-Funktion ermitteln:

$E = (50 - 0{,}5x) \cdot x = 50x - 0{,}5x^2$ und

$GE = 50 - x$ .

Die Gewinnmaximierungsbedingung des Monopolisten lautet GE = GK und damit:

$50 - x = 10 - 4x + x^2$ bzw.

$x^2 - 3x - 40 = 0$ .

Löst man diese quadratische Gleichung auf, so erhält man die Angebotsmenge x = 8. Setzt man nun diese Menge in die Preis-Absatz-Funktion ein, so folgt ein Preis von 46 Geldeinheiten, der über den Stückkosten von 19,5 Geldeinheiten bei dieser Menge liegt. Der Monopolist bietet folglich eine Menge von 8 Stück zu einem Preis von 46 Geldeinheiten an. Er erzielt damit pro Stück einen Gewinn von 26,5 Geldeinheiten. Diese ergeben sich aus der Differenz von Preis und Stückkosten bei der Menge x = 8.
Den Gesamtgewinn erhält man entweder durch Multiplikation der Menge 8 mit dem Stückgewinn von 26,5 Geldeinheiten oder durch Einsetzten der Menge 8 in die Gewinnfunktion:

$G = 8 \cdot 26{,}5 = 212$ bzw.

$$G(8) = E(8) - K(8) = 46 \cdot 8 - \left(10 \cdot 8 - 2 \cdot 8^2 + \frac{8^3}{3} + \frac{100}{3}\right) = 368 - 156 = 212.$$

Der Monopolist erzielt einen beachtlichen Gewinn von 212 Geldeinheiten.

# 4 Monopol versus homogenes Polypol

## 4.1 Vergleich der Marktgleichgewichte

Für eine Beurteilung der Wohlfahrts- und Verteilungswirkungen eines Monopols müssen die Marktergebnisse verschiedener Marktformen verglichen werden. Hierzu soll zunächst ein homogenes Polypol herangezogen werden. Bei dieser Marktform gibt es sehr viele Anbieter, die ein identisches Gut anbieten. Zudem liegt vollständige Markttransparenz vor.

Um solch einen Vergleich durchführen zu können, muss eine zentrale Annahme zur Kostensituation in den beiden Marktformen gemacht werden: Die Kostensituation im Monopol und im homogenen Polypol ist gleich. Dies bedeutet, dass die Grenzkostenfunktion des Monopolisten

der aggregierten Grenzkostenfunktion der Anbieter im homogenen Polypol entspricht (vgl. Herdzina, K., 2005, S. 171).

Bei vollständiger Konkurrenz ist nach (3b) der Grenzerlös gleich dem Preis des angebotenen Gutes. Die Anbieter können aufgrund ihrer geringen Größe den Marktpreis nicht beeinflussen und müssen ihn im Unterschied zum Monopolisten als Datum hinnehmen. Dadurch liegt das Gewinnmaximum an der Stelle, wo der Preis gleich den Grenzkosten ist. Der Preis muss zudem langfristig mindestens den Stückkosten entsprechen, da sonst dauerhafte Verluste erzielt werden:

(12) $p = GK \quad \text{mit} \quad p \geq k\,.$

So lange der Preis noch über dem Minimum der Stückkosten (= Betriebsoptimum) liegt, werden neue Anbieter auf den Markt strömen, da offensichtlich auf diesem Markt problemlos Gewinne erzielt werden können. Dieser Zustrom neuer Anbieter bewirkt bei vorhandener Nachfrage ein Sinken des Preises. Der Preis sinkt im Extremfall so weit, bis alle Anbieter in ihrem Betriebsoptimum und damit zu minimalen Stückkosten produzieren. Die Anbieter decken in dieser Situation genau ihre Kosten. Dabei sind eine angemessene Entlohnung des Unternehmers sowie eine landesübliche Verzinsung des Eigenkapitals als kalkulatorische Kosten berücksichtigt. Allerdings werden von den Anbietern im homogenen Polypol langfristig keine Extragewinne erzielt (vgl. Herdzina, K., 2005, S. 165).

Abb. 4 zeigt, dass das Marktgleichgewicht in einem homogenen Polypol bei einer größeren Menge $x_P$ und einem niedrigeren Preis $p_P$ liegt. Beides ist aus Sicht der Verbraucher positiv zu bewerten. Zusätzlich erkennt man, dass nur im Polypol langfristig im Betriebsoptimum produziert wird. Der Monopolist produziert eine geringere Menge zu höheren Stückkosten. Der Grund hierfür ist, dass er bei einer Ausweitung der Produktion die dann größere Angebotsmenge nur zu niedrigeren Preisen verkaufen könnte. Der Grenzerlös liegt dadurch unter den Grenzkosten, was zu einer Gewinnschmälerung führt.

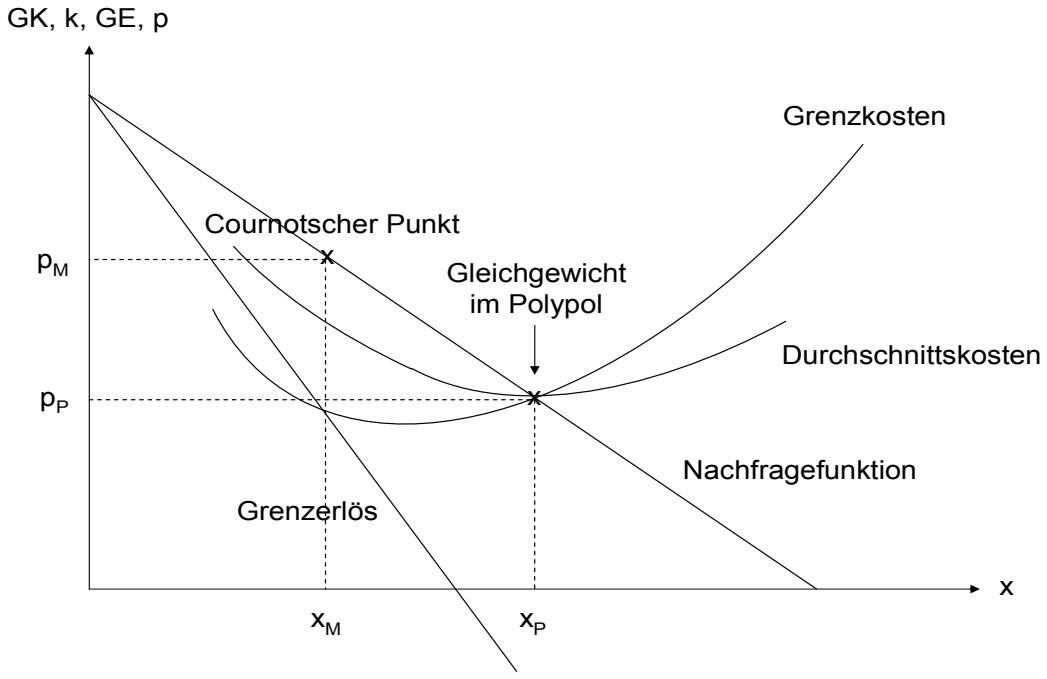



Abb. 4: Vergleich der Marktergebnisse im Monopol und homogenen Polypol

Obwohl der Monopolist zu höheren Stückkosten produziert, erzielt er im Unterschied zu den Anbietern im homogenen Polypol einen Extragewinn. Der Monopolpreis liegt über den Stückkosten. Die Differenz ist ein zusätzlicher Stückgewinn. Dagegen produzieren die Anbieter in der Konkurrenzsituation durch den ständigen Markteintritt neuer Wettbewerber langfristig zu Stückkosten, die genau dem Preis entsprechen. Ein Extragewinn wird nicht erzielt (vgl. Herdzina, K., 2005, S. 173).

## 4.2 Konsumenten- und Produzentenrente

Der bisherige Vergleich erlaubt nur die Aussage, dass die Verbraucher in einem Monopol schlechter gestellt werden als in einem homogenen Polypol. Gleichzeitig wird aber das Unternehmen bei einem Monopol besser gestellt. Es findet also eine Umverteilung statt, die nur über ein Werturteil beurteilt werden kann (vgl. Varian, H.R., 2007, S. 506f.). Es lässt sich nun jedoch zeigen, dass Monopole auch aus Effizienzüberlegungen problematisch sind.

Hierzu muss auf die Konsumenten- und die Produzentenrente zurückgegriffen werden. Betrachtet man den Markt für ein beliebiges Gut, so stellt sich bei frei beweglichem Preis im Marktgleichgewicht ein Ausgleich von Angebot und Nachfrage ein. Viele Nachfrager wären aber bereit gewesen, für dieses Gut auch einen höheren Preis zu bezahlen, da ihnen dieses Gut einen sehr hohen Nutzen stiftet. Die Differenz zwischen dem Preis, den sie zu zahlen bereit wären, und dem Marktpreis bezeichnet man als Konsumentenrente. Für die Anbieter gilt eine ähnliche Überlegung. Möglicherweise wäre ein Teil der Anbieter bereit gewesen, auch zu einem niedrigeren Preis anzubieten. Sie erhalten aber den höheren Marktpreis und erzielen dadurch eine Produzentenrente.

In Abb. 5 sind zwei unterschiedliche Marktgleichgewichte eingezeichnet. Herrscht auf einem Markt eine Konkurrenzsituation, so ergibt sich ein Gleichgewicht im Schnittpunkt von Angebots- und Nachfragefunktion. Die Konsumentenrente $KR_P$ setzt sich in diesem Fall aus den Flächen A, B und D zusammen. Die Produzentenrente $PR_P$ umfasst im Polypol die Flächen C und E. In der Summe umfassen die Renten somit die Dreiecksfläche, die von Angebots- und Nachfragefunktion eingeschlossen wird.

Der Monopolist wählt dagegen auf der Nachfragefunktion die für ihn gewinnmaximale Preis-Mengen-Kombination aus. Die Darstellungen zum Cournotschen Punkt haben gezeigt, dass die ausgewählte Kombination niedrigere Mengen und höhere Preise bedeutet. Die Konsumentenrente $KR_M$ sinkt in diesem Fall auf die Dreiecksfläche A. Die Produzentenrente $PR_M$ besteht bei gleicher Kostensituation nun aus den Flächen B und C.

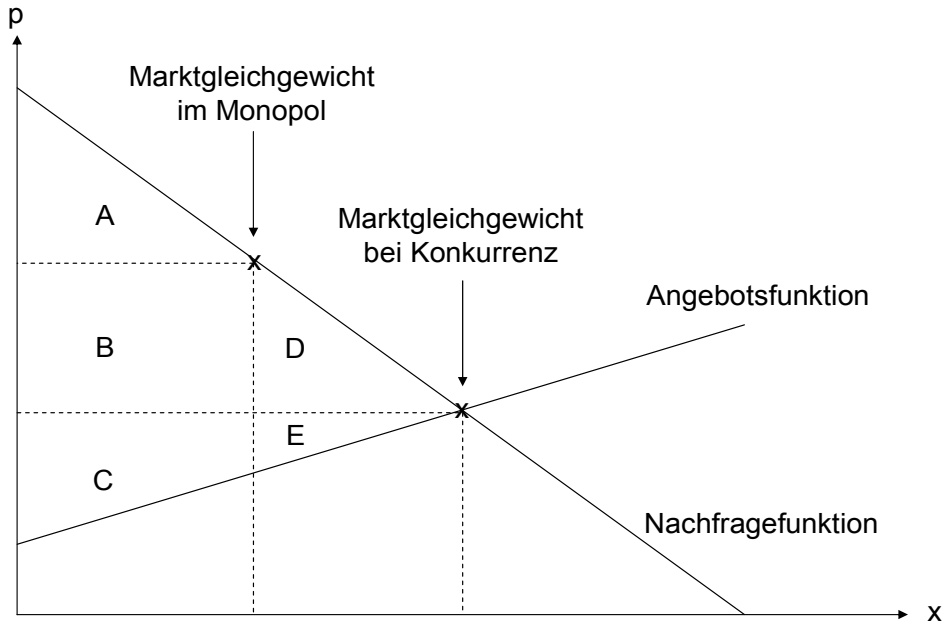



Abb. 5: Entwicklung von Konsumenten- und Produzentenrente

Ein Vergleich der Gesamtrenten zeigt, dass im Monopolfall Konsumenten- und Produzentenrente insgesamt um die Flächen D und E niedriger sind. Es hat also nicht nur eine Umverteilung zugunsten des Monopolisten stattgefunden, sondern es kommt auch zu Wohlfahrtsverlusten. Ein Monopol ist daher als eine ineffiziente Marktstruktur einzustufen (vgl. Varian, H. R., 2007, S. 508ff.):

(13) $KR_P + PR_P = (A + B + D) + (C+E) > KR_M + PR_M = A + (B + C)$ .

Der Vergleich verschiedener Marktstrukturen ist allerdings nicht unproblematisch. Vor allem die Annahme, dass die Kostensituation des Monopolisten der aggregierten Kostensituation vieler kleiner Anbieter gleicht, ist auf zahlreichen Märkten realitätsfern. Existieren steigende Skalenerträge, so können größere Anbieter die gleiche Menge günstiger produzieren als viele kleine Anbieter. Im Extremfall können diese Massenproduktionsvorteile sogar in einem natürlichen Monopol münden (vgl. Herdzina, K., 2005, S. 188). Solche Strukturen spielen insbesondere bei netzgebundenen Dienstleistungen mit einem sehr hohen Fixkostenanteil eine wichtige Rolle. Beispiele hierfür sind die Bereiche Telekommunikation, Bahn und Energieversorgung. In solchen Fällen sind Marktstrukturen mit vielen kleinen Anbieter nicht zu verwirklichen und aus Kostengründen volkswirtschaftlich auch nicht wünschenswert. Allerdings kann auch in diesen Branchen durch die Trennung von Netzbetrieb und Dienstleistungsbetrieb wenigstens bei den Dienstleistungen ein gewisses Maß an Wettbewerb erreicht werden.

## 5 Oligopole und Kollektivmonopole

Auf vielen Gütermärkten liegen weder Monopole noch (homogene) Polypole vor. Stattdessen findet man einige wenige Anbieter, die um die Nachfrage konkurrieren. Man denke bspw. an die Mineralölkonzerne oder an die Netzbetreiber auf dem Mobilfunkmarkt. Im Rahmen der Spieltheorie werden zahlreiche Lösungen für diese Oligopole untersucht. Dabei zeigen sich sehr unterschiedliche Marktergebnisse, die jeweils von bestimmten Annahmen abhängen (vgl. Bartmann, H./ Busch, A. A./Schwaab, J. A., 1999, S. 161ff.).

In vielen Fällen bietet es sich für die Anbieter an, die oft unabwägbaren Konsequenzen des Wettbewerbs durch Absprachen auszuschließen. Die Anbieter verhalten sich bei solch einem Kartell gemeinsam so, wie sich ein Monopolist verhalten würde. Man spricht daher auch von einem Kollektivmonopol (vgl. Frantzke, A., 2004, S. 240). Durch dieses gemeinsame Verhalten können die Oligopolisten den gleichen Extragewinn erzielen wie ein Monopolist. Dieser Gewinn muss dann noch auf die verschiedenen Anbieter aufgeteilt werden.

Kollektivmonopole werden begünstigt, wenn nur noch relativ wenige Anbieter auf einem Markt vorhanden sind. Eine Absprache ist zu fünft einfacher als eine Absprache zwischen zwanzig Unternehmen. Ist das hergestellte Gut sehr homogen, erleichtert dies ebenfalls die Kartellbildung. So sind Absprachen in der Zementindustrie eher zu erwarten als zwischen verschiedenen Automobilherstellern.

## 6 Notwendigkeit einer Wettbewerbspolitik

Die bisherigen Überlegungen zeigen, dass sowohl Monopole als auch Oligopole mit sehr wenigen Anbietern in vielen Fällen als wohlfahrtsmindernde Marktkonstellationen anzusehen sind. Das Grundproblem liegt aber nicht darin, dass der Monopolist „böser" ist als der Anbieter bei vollständiger Konkurrenz. Beide versuchen lediglich ihren Gewinn zu maximieren. Dieses Verhalten ist nicht nur typisch für Unternehmen in einer Marktwirtschaft, sondern weitgehend auch erwünscht. Das Problem ist also nicht die Verhaltensweise an sich, sondern die Marktstruktur, die automatisch zu unbefriedigenden Marktergebnissen führt (vgl. Mankiw, N.G., 2004, S. 354f.).

Daraus ergeben sich vielfältige Konsequenzen für die Wettbewerbspolitik. Monopole und enge Oligopole sollten möglichst durch Fusionskontrollen und Kartellverbote verhindert werden. In den Fällen, in denen Marktstrukturen mit wenigen Anbietern aus Kostengründen sinnvoll sind, muss eine Missbrauchsaufsicht sicherstellen, dass Monopolisten oder Oligopolisten ihre Marktmacht nicht unangemessen ausnutzen. Sowohl das deutsche Gesetz gegen Wettbewerbsbeschränkungen (GWB) als auch das europäische Wettbewerbsrecht tragen diesen Gedanken Rechnung.

## Literatur

Bartmann, H.; Busch, A. A.; Schwaab, J. A. (1999): Preis- und Wettbewerbstheorie, 6. Aufl., St. Gallen 1999

Cournot, A. A. (1924): Untersuchungen über die mathematischen Grundlagen der Theorie des Reichtums (übersetzt von W. G. Waffenschmidt), Jena 1924

Frantzke, A. (2004): Grundlagen der Volkswirtschaftslehre – Mikroökonomische Theorie und Aufgaben des Staates in der Marktwirtschaft, 2. Aufl. , Stuttgart 2004

Herdzina, K. (2005): Einführung in die Mikroökonomik, 10. Aufl., München 2005

Mankiw, N. G. (2004): Grundzüge der Volkswirtschaftslehre, 3. Aufl., Stuttgart 2004

Petersen, T. (2008): Preisbildung auf Monopolmärkten, in: WISU – Das Wirtschaftsstudium, 37. Jg., Heft 1/2008, S. 67-70.

Varian, H. R. (2007): Grundzüge der Mikroökonomik, 7. Aufl., München 2007

Moore, H. L.; Waffenschmidt, W.G. (1965): Antoine Augustin Cournot (1801-1877), in: Recktenwald, H.C., Lebensbilder großer Nationalökonomen – Einführung in die Geschichte der Politischen Ökonomie, Köln/Berlin 1965, S. 231-246

Zimmermann, L. J. R.; Marcon, H. (1989): Antoine Augustin Cournot (1801-1877), in: Starbatty, J., Klassiker des ökonomischen Denkens, 1. Band, München 1989, S. 245-265

# Methoden des Operations Research und ihre Anwendung in der Betriebswirtschaftslehre[1]

*von Ralf Gerhards*

## 1 Gegenstand, Methodik und Techniken des Operations Research

Es gibt verschiedene deutsche Bezeichnungen für den Begriff des Operations Research (OR). Am treffendsten scheint Operations Research mit dem Begriff „Planungsforschung" umschrieben zu sein. Demnach hat OR die Aufgabe, Methoden bereitzustellen, um Planungsaufgaben in der betrieblichen Realität bewältigen zu können. Verfahren des OR werden im betriebswirtschaftlichen Umfeld vor allem in der Investitionsprogrammplanung, in der Produktionsprogrammplanung und im Transport- und Lagerwesen eingesetzt.

Die Verfahren des OR sind Problemlösungsverfahren und dienen folglich der Entscheidungsunterstützung für die unternehmerische Exekutive. Die Entscheidungen selbst sind dann vom Management zu fällen. In diese Entscheidungen fließen in der Regel nicht nur die Ergebnisse des OR ein. Vielmehr können auch weitere entscheidungsrelevante Parameter einfließen, etwa unternehmenspolitische Impulse, die dann die schlussendliche Entscheidung begründen. In einer ökonomischen Welt, in der nur rational Handelnde (im Sinne eines Homo Oeconomicus) agieren, wären die Ergebnisse des OR gleichzeitig auch die Entscheidung zur Lösung des Problems.

Die Problemlösungsverfahren des OR sind dadurch gekennzeichnet, dass sie versuchen, Optimallösungen für die gestellten Probleme quantitativ aufzuzeigen. Dies erfordert den Einbezug ökonomischer, mathematischer und statistischer Modelle und Verfahren, in denen das Risiko und die Unsicherheit, die Planungsaufgaben stets umgeben, in die Betrachtung einbezogen werden.

Ausgangspunkt jedweder Verfahren des OR ist die Untersuchung des Planungsprozesses und seiner innewohnenden Prämissen, Risiken und Unsicherheiten zwecks Entscheidungsvorbereitung. Diese Untersuchungen können in strukturierter Form im Rahmen eines Untersuchungsplans sequentiell als Abfolge abgebildet werden:

[1] Vergleiche einführend *Schneeweiß, Christoph*: Operations Research. In: Handwörterbuch der Betriebswirtschaft, Teilband 2, 5. Auflage, Spalten 2940-2953. Stuttgart 1993 sowie *Bartels, Hans*: Optimierung, lineare. In: Handwörterbuch der Betriebswirtschaft, Teilband 2, 5. Auflage, Spalten 2953-2968. Stuttgart 1993.

1. Formulierung des Entscheidungsproblems aus der realen Welt, also Ableitung des konkreten Handlungsbedarfs
2. Transformation des Problems aus der realen Welt in ein Modell. Ein Modell ist eine homomorphe Abbildung der Realität und damit eine Abstraktion der Realität. Bei der Modellbildung sind folgende Ablaufschritte durchzuführen:
   a. Identifikation (verbal und formal) der Parameter und der Variablen, die im Entscheidungsproblem involviert sind
   b. Beschränkung auf jene Parameter (Konstanten, Variablen mit genau einem Wert) und Variablen, welche die wesentlichen Einflüsse im Entscheidungsproblem darstellen. Durch die Fokussierung auf die wesentlichen Elemente soll das Modell überschaubar und handhabbar gestaltet werden. Die Variablen im Entscheidungsproblem sind zu kategorisieren in solche, die beeinflussbar und damit kontrollierbar sind sowie in jene, die nicht beeinflussbar und damit nicht kontrollierbar sind.
   c. Die Beziehungen der Variablen untereinander sind transparent zu machen. Dabei spielt bekanntes Wissen über die Abhängigkeiten eine genauso große Rolle wie die Intuition. Diese wird gebraucht, wenn es kein gesetztes Wissen über die Variablenbeziehungen gibt und man daher auf „den gesunden Menschenverstand" zurückgreifen muss.
   d. Für die nicht kontrollierbaren Variablen und für Beziehungen zwischen Variablen, über die es keine theoretischen Erkenntnisse gibt, gilt, dass für ihr Verhalten im Modell Annahmen und Prognosen aufgestellt werden müssen. Die Art der Beziehungen zwischen Variablen kann über Korrelationen statistisch ermittelt werden.
   e. Das Modell ist zunächst verbal zu beschreiben und schließlich in ein System symbolischer Beziehungen zu übersetzen. Damit erhält das Modell eine formale (mathematische) Struktur. Ein in eine formale Struktur überführtes Modell wird als *formales Entscheidungsmodell* bezeichnet. Damit ist letztlich der Übergang von der (homomorphen) Realebene zur Formalebene erfolgt. Die getroffenen Annahmen über Variablen beziehungsweise über die Beziehungen zwischen Variablen werden über Nebenbedingungen im formalen Entscheidungsmodell berücksichtigt.
3. Nachdem das Entscheidungsmodell vorliegt, kann das in einem Modell gefasste Entscheidungsproblem gelöst werden. Man bezeichnet diesen Schritt auch als *Entscheidungsfindung*. Eine Gleichsetzung des OR mit der Entscheidungsfindung wäre falsch. Eine erfolgreiche Lösung des Modells setzt die Kenntnis des gesamten Planungsprozesses voraus. OR ist demnach als ganzheitlicher Prozess der Planungsforschung aufzufassen. Die Modellösung erfolgt über die Anwendung quantitative Methoden, die zu einer (rechnerischen) Optimallösung führen und damit dem Entscheidungsträger Impulse zur Entscheidungsfindung bereitstellen. Die Optimierung ist dabei stets eine Maximierung oder Minimierung einer Zielfunktion unter Beachtung der im Modell aufgezeigten Beziehungen (Nebenbedingungen).
4. Schließlich sind die Ergebnisse aus der Entscheidungsfindung zu realisieren, also zu implementieren und zu kontrollieren. Die Aufgabe der Kontrolle ist hierbei, die im Modell getroffenen Annahmen zu überprüfen und, gegebenenfalls, Modellrevisionen zu initiieren. Die Implementierung der Modellergebnisse sowie die Kontrolle sind nicht mehr zentrale Aufgabe des OR. Vielmehr handelt es sich hierbei um datenverarbeitungstechnische Aufgaben (Implementierung) und um Führungsaufgaben (Kontrolle).

Die „Toolbox" des OR enthält verschiedene Optimierungstechniken, die jeweils für spezifische Planungsprobleme angewendet werden können:

- *Lineare Optimierung (Lineare Programmierung, linear programming)*
  Die lineare Programmierung löst solche Planungsprobleme, deren Struktur sich als System linearer Gleichungen und linearer Bedingungen darstellen lässt. In der linearen Programmierung werden drei Funktionen beschrieben, nämlich die Zielfunktion (Ma-

ximierungsproblem oder Minimierungsproblem), die Nebenbedingungen sowie die Nicht-Negativitätsbedingung (alle Variablen sind positiv oder Null).

- *Warteschlangentheorie (queueing theory)*
  Warum bilden sich Warteschlangen an Bankschaltern? Warum stauen sich Aufträge vor Maschinen? Die Warteschlangentheorie versucht Antworten darauf zu geben, warum es zu Engpässen kommt. Die Optimierungsintension liegt darin begründet, dass ein Optimum gefunden werden muss zwischen den Unterhaltungskosten der Bedienungseinrichtung (es könnten im Bankbeispiel zusätzliche Schalter angeboten werden) und den Wartekosten der abzufertigenden Objekte. Die Warteschlangentheorie ermittelt zum Beispiel die optimale Anzahl an Beschäftigten einer Materialausgabestelle. Optimiert (hier: minimiert) wird die Summe aus den Personalkosten der Schalterbeschäftigten und den Leerkosten der wartenden Beschäftigten. In den Betrachtungshorizont fließen insbesondere die durchschnittliche Ankunfts- und Abfertigungsrate, die mittlere Schlangenlänge sowie die durchschnittliche Wartezeit ein.
- *Lagerhaltungsmodelle (inventory models)*
  Lagerhaltungsmodelle versuchen einen Ausgleich zu schaffen zwischen den Kosten der Lagerhaltung und den Kostenkosten der Bestellung. Diese Kosten verhalten sich gegenläufig zueinander. Hohe Lagerkapazitäten (und damit hohe Lagerkosten) ermöglichen die Reduzierung der Bestellkosten etwa durch das Ausnutzen von Mengenrabatten. Umgekehrt führen kleinere Bestellmengen einerseits zu höheren Bestellkosten und andererseits zu geringeren Lagerkosten, da weniger Lagerkapazitäten benötigt werden. Die Materialwirtschaft errechnet zur Schaffung des Kostenausgleichs die optimale Bestellmenge.
- *Netzplantechnik (network techniques)*
  Die OR-Methode der Netzplantechnik dient der Planung und Steuerung insbesondere von Projekten. Alle Vorgänge eines Projektes (Projektphasen) werden in ihren technologisch oder wirtschaftlich bedingten Reihenfolgenbeziehungen festgesetzt. Ferner fließen in einen Netzplan Ereignisse, Vorgangsdauern, Ressourceneinsatz und zeitliche Abhängigkeiten (früheste und späteste Beginn- und Endzeitpunkte; kürzeste Projektdauer) ein. Neben diesen Parametern, die Restriktionen eines Projektes darstellen, fließen vor allem auch Budgetbeschränkungen und Kostenvorgaben ein, die simultan zu den anderen Bedingungen ganzheitlich optimiert werden müssen.
  Ein Netzplan ist zunächst ein graphentheoretisches Modell, in denen die Vorgänge Knoten dargestellt werden und Pfeile die Reihenfolgebeziehungen widerspiegeln.
- *Ersatzmodelle (replacement models)*
  Ersatzmodelle befassen sich mit der Determinierung des Lebenszyklus von Gegenständen und damit mit deren (kosten-) optimalen Restitution. Gegenstände können entweder sofort ersetzt werden bei jedem Ausfall oder es werden mehrere (gegebenenfalls alle) Gegenstände nach einer bestimmten Zeit auf einmal erneuert. Das Aufzeigen der optimalen Ersatzstrategie ist Ziel der Ersatzmodelle. Beim Gruppenersatz sind nämlich regelmäßig die Ersatzkosten pro Stück niedriger verglichen mit dem Einzelersatz. Gleichzeitig entstehen Kosten aufgrund des Ersatzes grundsätzlich noch funktionsfähiger Gegenstände. Die Optimierung besteht hier in der Ermittlung des Kostenminimums.
- *Dynamische Programmierung (dynamic programming)*
  Die dynamische Programmierung beinhaltet Methoden zur Optimierung mehrstufiger Prozesse, bei denen die Entscheidung auf einer Stufe jene der jeweils nächsten Stufe beeinflusst. Insbesondere Produktionsplanungsprobleme (mehrstufige Produktion) bedienen sich dieser Verfahren.

- *Simulationsverfahren (simulation)*
  Wo mathematische Verfahren keine Modellösungen generieren können, versuchen Simulationsverfahren einen Beitrag zur Entscheidungsfindung zu liefern. Bei den Simulationsverfahren handelt es sich um experimentelle Methoden, die durch Versuche (Probieren) eine Näherungslösung anstreben.
- *Spieltheorie (game theory)*
  Die Spieltheorie will das Verhalten rational (den Gesetzen der Logik folgender[2]) agierender Spieler mathematisch analysieren, indem Strategien für Entscheidungen entwickelt werden sowie Handlungsempfehlungen in Konfliktsituationen als auch in Situationen mit Unsicherheit gegeben werden. Gerade die Spieltheorie ist aktuell in der betriebswirtschaftlichen Forschung durch grundlegende neue Beiträge bereichert worden. Der Nobelpreis für Wirtschaft für das Jahr 2005 wurde für Forschungsergebnisse zur Spieltheorie vergeben. Robert Aumann und Thomas Schelling erhielten im Jahr 2005 den Nobelpreis für „ihre grundlegenden Beiträge zur Spieltheorie und zum besseren Verständnis von Konflikt und Kooperation". Bereits 1994 wurde der Nobelpreis für Wirtschaft für die „grundlegende Analyse des Gleichgewichts in nicht-kooperativer Spieltheorie" an die Amerikaner John Charles Harsanyi und John Forbes Nash sowie an den Deutschen Reinhard Selten vergeben.
- *präskriptive Entscheidungsmodelle (Entscheidungsorientierte Betriebswirtschaftslehre)*
  Die präskriptive Entscheidungstheorie versucht unter der Annahme rationaler Grundeinstellung der Entscheidungsträger und insbesondere in Abhängigkeit von deren Risikoeinstellung logische Entscheidungsregeln aufzusetzen. Dabei wird das Entscheidungsproblem in eine Matrix aus Umweltzuständen und Handlungsalternativen transformiert. Der jeweilige Schnittpunkt aus Umweltzustand und Handlungsalternative ist dann ein Ergebnis. Die Entscheidungsregeln wählen dann logisch die Lösung aus der gesamten Ergebnismatrix aus. Man könnte sagen, sie geben eine Lösung eines Entscheidungsproblems in Form einer Empfehlung der Art „Du sollst" vor.

# 2 Einsatz des Operations Research in der Betriebswirtschaftslehre

Optimierungsprobleme in der Betriebswirtschaftslehre lassen sich bereits aus dem Wirtschaftlichkeitsgrundsatz ableiten, der die zentrale Rechtfertigung der Betriebswirtschaftslehre insgesamt darstellt. Demnach zwingt die Knappheit der Ressourcen (Güter, Zeit) und die grundsätzlich unbegrenzten menschlichen Bedürfnisse zu wirtschaftlichem Verhalten. Die Knappheit beziehungsweise die Begrenztheit der Ressourcen hat nämlich zur Folge, dass sie nicht umsonst zu haben sind. Also ist mit den Ressourcen zu haushalten. Das Ziel der optimalen Bedürfnisbefriedigung ist unter der zentralen Nebenbedingung der Knappheit der Ressourcen *das* zentrale Optimierungsproblem in der Betriebswirtschaftslehre. Unternehmen, also Betriebe der Fremdbedarfsdeckung, versuchen in einer Marktwirtschaft, unter gegebenen (begrenzten) Produktionsfaktoren (Ressourcen) den Gewinn zu optimieren. Private Haushalte, also Betriebe der Eigenbedarfsdeckung, optimieren ihre Bedürfnisbefriedigung unter der Nebenbedingung begrenzter finanzieller Budgets.

Sämtliche Funktionsbereiche der Betriebswirtschaftslehre sind, zumindest partiell, einer quantitativen Analyse mittels der Verfahren des OR zugänglich. In allen Funktionsbereichen, in manchen mehr, in manchen weniger, pflanzt sich das zentrale Optimierungsproblem, also die Notwendigkeit zum Wirtschaften, fort.

[2] Gefühlsgeleitete Entscheidungen werden in der Spieltheorie ignoriert.

Für die Anwendung der Methoden des OR besonders geeignet sind die folgenden Einsatzbereiche:

- Logistik (Transportprobleme)
- Produktionsplanung (bei knappen Rohstoffen)
- Projektmanagement

Wie bei anderen wissenschaftlichen Disziplinen auch haben sich im OR zwei Richtungen herausgebildet. In beiden Richtungen steht die mathematische Optimierung im Mittelpunkt und hierbei vor allem die Methode der linearen Programmierung sowie die Netzplantechnik. Die verfahrensorientierte Richtung des OR versucht, die mathematischen Verfahren zur Lösung der Modelle weiter zu entwickeln. Insofern wird dieser Zweig des OR häufig in der angewandten Mathematik, aber auch im Bereich des Wirtschaftsingenieurwesens und in der Wirtschaftsinformatik thematisiert.

Die anwendungsorientierte Richtung des OR hat zum Ziel, betriebswirtschaftliche Anwendungen dem OR zu öffnen und dabei die betriebswirtschaftliche Problemstellung zu analysieren und zu fixieren. Das betriebswirtschaftliche Problem wird dann mittels mathematischer Verfahren zu lösen versucht. Kenntnisse auf diesem Gebiet werden in der Betriebswirtschaftslehre vermittelt.

Obwohl beide Bereiche, also sowohl das Aufstellen des betriebswirtschaftlich fundierten Problems als auch dessen mathematische Lösung, zwei Seiten derselben Medaille sind, hat sich eine „Arbeitsteilung" als pragmatisch erwiesen. Die anwendungsorientierte Richtung des OR liefert den betriebswirtschaftlichen Input und die verfahrensorientierte Richtung den mathematisch abgeleiteten Output, also die Modellösung.

In einer Lehrveranstaltung der anwendungsorientierten Richtung steht wohl das betriebswirtschaftliche Planungsproblem im Vordergrund. Kenntnisse zumindest in den zentralen Methoden des OR (vor allem lineare Programmierung und Netzplantechnik) sind aber auch für Studierende der Betriebswirtschaftslehre unabdingbar, da sie zum gesamten Verständnis des Operations Research unabdingbar vonnöten sind.

# 3 Ausgewählte Methoden des Operations Research

## 3.1. Präskriptive Entscheidungsmodelle

### 3.1.1 Grundlagen und Überblick

Das Grundmodell der Entscheidungslogik bildet eine Matrix aus mit den Handlungsalternativen ($a_i$) und den möglichen Umweltzuständen ($z_i$), die in ihren jeweiligen Schnittpunkten zu Ergebnissen ($e_i$) führen:

| Umweltzustände→<br>Handlungsalternativen<br>↓ | $z_1$ | $z_2$ | $z_3$ | $z_n$ |
|---|---|---|---|---|
| $a_1$ | $e_{1,1}$ | $e_{1,2}$ | $e_{1,3}$ | $e_{1,n}$ |
| $a_2$ | $e_{2,1}$ | … | … | … |
| $a_3$ | … | … | … | … |
| $a_m$ | … | … | … | $e_{m,n}$ |

Durch die Anwendung entscheidungslogischer Verfahren soll das Eintreten von Umweltzuständen rational erfasst werden. Im engeren Sinne geht es bei der präskriptiven Entscheidungstheorie um die rationale Erfassung von Unsicherheit, die, in mehr oder weniger starker Form aller Planungen charakterisiert.

Beispiel:
Bei den Handlungsalternativen möge es sich um verschiedene geplante absatzfördernde Marketingmaßnahmen handeln. Es kann sich dabei um Werbung im Fernsehen handeln, um eine Unternehmens- beziehungsweise Produktpräsentation in der Innenstadt oder um die Werbung in Zeitungen und Zeitschriften. Die Umweltzustände können verschiedene Annahmen über die Entwicklung des allgemeinen Lohnniveaus reflektieren. Die Steigung des Lohnniveaus kann über der Inflation liegen, auf Inflationsniveau oder darunter. Es wird unterstellt, dass sich die Entwicklung des Lohnniveaus direkt auf die Absatzmenge handelt. Die Ergebnisse sollen die erwarteten Umsätze in Millionen Euro spiegeln.

Sind Wahrscheinlichkeiten über das Eintreten von Umweltzuständen bekannt, so spricht man von „Entscheidungen unter Risiko". Die Ergebnisse und damit die Zielwerte hängen von dem Eintreten eines Umweltzustandes ab. Hinsichtlich des Eintretens des Umweltzustandes existieren Wahrscheinlichkeiten.

Liegen solche Wahrscheinlichkeiten nicht vor (entweder weil sie nicht ermittelbar sind, oder weil ihre Ermittlung mit zu hohem Aufwand verbunden ist oder weil man keine Wahrscheinlichkeiten in die Entscheidungsfindung einfließen lassen möchte), so spricht man von „Entscheidung unter Ungewissheit". Auch hier hängen die Ergebnisse und damit die Zielwerte von dem Eintreten eines Umweltzustandes ab. Zur Eintrittswahrscheinlichkeit des Umweltzustands liegen keine Informationen vor.

### 3.1.2 Ausgewählte Entscheidungsregeln im Überblick

Ausgangspunkt für die Anwendung von Entscheidungsregeln sei folgende Matrix unter Zugrundelegung des Beispiels aus Abschnitt 3.1.1:

| Umweltzustände→<br>Handlungsalternativen<br>↓ | Lohnsteigerung überkompensiert die Inflationsrate | Lohnsteigerung kompensiert die Inflationsrate | Lohnsteigerung liegt unter der Inflationsrate |
|---|---|---|---|
| Spot im Fernsehen | 70 | 80 | 260 |
| Präsentationsstand | 140 | 100 | 90 |
| Werbeannonce in der überregionalen Presse | 120 | 95 | 105 |

Bei der *Bayes-Regel*, die eine Entscheidungsregel unter Risiko darstellt, wird der maximale Erwartungswert gesucht und jene Handlungsalternative ausgewählt, welche den höchsten Erwartungswert hat[3]. Der Erwartungswert reflektiert die mit den anzugebenden Wahrscheinlichkeiten gewichteten Zielerreichungsgrade.

Die angenommenen Wahrscheinlichkeiten betragen für das Eintreten des Umweltzustands $z_1$ (Lohnsteigerung überkompensiert die Inflationsrate) 30% und jene für die Umweltzustände $z_2$ und $z_3$ 50% beziehungsweise 20%.

Die Bayes-Regel ermittelt für Handlungsalternative $a_1$ einen Zielerreichungswert von (70*0,3)+(80*0,5)+(260*0,2) = 113. Für $a_2$ und $a_3$ ergeben sich die Werte 110 und 104,5.

Wenn man die Anwendung der Bayes-Regel beurteilt, so fällt zunächst die einfache Handhabung dieser Regel auf. Nachteilig ist allerdings, dass sie ein risikoneutrales Verhalten des Entscheidungsträgers voraussetzt. Es wird nämlich ein einem unwahrscheinlichen Umweltzustand prognostizierten Ergebnis ein einem wahrscheinlichen Umweltzustand zuzurechnendes niedriges Ergebnis einer Handlungsalternative gleichgestellt. Ein Ergebnis von 50 bei einer Wahrscheinlichkeit von 30% ist gleich einem Ergebnis von 30 bei einer Eintrittswahrscheinlichkeit von 50%.

Zu den Entscheidungsregeln unter Ungewissheit zählen unter anderem die MaxiMin-Regel sowie die Savage-Niehaus-Regel.

Bei der *MaxiMin-Regel* werden zunächst die Zeilenminima ermittelt. In einem zweiten Schritt wird das Maximum der Minima ermittelt. Diese pessimistische, risikoaverse Entscheidungsregel geht vom Eintreffen des „worst case" bei jeder Handlungsalternative aus und wählt jene Handlungsalternative aus, bei der das schlechteste Ergebnis noch das höchste im Vergleich zu den anderen Handlungsalternativen darstellt. Im Beispiel betragen die Zeilenminima 70, 90 beziehungsweise 95. Demnach würde man bei Anwendung der MaxiMin-Regel Handlungsalternative $a_3$ wählen.

Die *Savage-Niehans-Regel* (Entscheiden nach kleinstem Bedauern) ermittelt zunächst die Spaltenmaxima. Dann werden die Ergebniswerte vom jeweiligen Spaltenmaximum subtrahiert. Schließlich wird das Zeilenmaximum ermittelt. Gewählt wird jene Handlungsalternative, die das kleinste Zeilenmaximum hat und damit das „kleinste Bedauern" bewirkt. Das Entscheidungstableau im Beispiel wird also in eine Matrix maximaler Bedauernswerte transformiert:

| Umweltzustände→<br>Handlungsalternativen<br>↓ | Lohnsteigerung überkompensiert die Inflationsrate | Lohnsteigerung kompensiert die Inflationsrate | Lohnsteigerung liegt unter der Inflationsrate |
|---|---|---|---|
| Spot im Fernsehen | 70 | 20 | 0 |
| Präsentationsstand | 0 | 0 | 170 |
| Werbeannonce in der überregionalen Presse | 20 | 5 | 155 |

Nach der Savage-Niehans-Regel wird Handlungsalternative $a_1$ (Spot im Fernsehen) gewählt. Bei ihr ist das Zeilenmaximum verglichen mit $a_2$ und $a_3$ am kleinsten.

[3] Im Ausgangsbeispiel werden die Umsätze geplant. Insofern wird ein maximales Ergebnis angestrebt. Wenn es sich bei den Ergebnissen etwa um Kosten handeln würde, so würde ein minimales Ergebnis angestrebt.

### 3.1.3 Würdigung präskriptiver Entscheidungsregeln

An den präskriptiven Entscheidungsregeln lassen sich verschiedene Kritikpunkte festmachen:

- Wie bei allen Verfahren des Operations Research wird ein rationales Verhalten der Entscheidungsträger vorausgesetzt.
- Informationsprobleme, nämlich hinsichtlich der Bestimmung der Handlungsalternativen sowie der Umweltzustände als auch der Ergebnisse werden ignoriert. Es werden damit vollständige Informationen unterstellt, die man nur aus vollkommenen Märkten kennt. Bekanntermaßen sind vollkommene Märkte mehr als theoretisches Konstrukt aufzufassen.
- Entscheidungsaspekte werden isoliert voneinander betrachtet. Es gibt demnach keine Interdependenzen zwischen den Handlungsalternativen. Auch dies ist, mit der Praxis gespiegelt, nicht per se anzunehmen.
- Die subjektive Risikoeinstellung der Entscheidungsträger muss bekannt sein. Nur dann können „passende" Entscheidungsregeln ausgewählt werden. Eine diffuse Risikoeinstellung könnte dazu führen, dass verschiedene Entscheidungsregeln, jene für risikobereite, jene für risikoneutrale und jene für risikoaverse Entscheidungsträger zur Ableitung einer Entscheidungsempfehlung herangezogen werden. Ergeben alternative Entscheidungsregeln, jeweils bezogen auf das gleiche Entscheidungstableau, unterschiedliche Entscheidungsempfehlungen, so kann eine schlussendliche Entscheidungsvorlage nicht geliefert werden.
- Die Entscheidungsregeln geben keine Entscheidungshilfe in solchen Fällen, in denen mehrere Handlungsalternativen das beste Ergebnis haben.
- Der Modellcharakter sorgt dafür, dass die Realität, die dem Entscheidungsproblem zugrunde liegt, nur mehr oder weniger ähnlicher Form wiedergegeben wird.

## 3.2. Entscheidungsbaumanalyse mittels der „Bayes Formel"

### 3.2.1. Grundlagen und Überblick

Bei der Entscheidungsbaumanalyse handelt es sich um eine formale Entscheidungsanalyse. Es ist ein weit verbreitetes, entscheidungstheoretisches Verfahren zur Unterstützung der Entscheidungsfindung. Der Entscheidungsbaum wird in aller Regel graphisch dargestellt und kann als gerichteter Graph aufgefasst werden, der verschiedene Knoten umfasst. Bei den Knoten handelt es sich entweder um Entscheidungsknoten oder um Zustandsknoten. Zustandsknoten repräsentieren im Entscheidungsbaum grundsätzlich mögliche Umweltzustände dar. Sie werden im Entscheidungsbaum in Kreisform dargestellt. Entscheidungsknoten stellen mögliche Entscheidungsoptionen dar, die sich einander ausschließen können. Je nach Entscheidung ergeben sich n verschiedene Zustände. Im Entscheidungsbaum werden die Entscheidungsknoten als Quadrat abgebildet.

Grundsätzlich ist also die Entscheidungsbaumanalyse eine mehrstufige Hilfe, unter der Annahme bestimmter Umweltzustände Erwartungen über die (quantitativen) Entscheidungskonsequenzen abzubilden. Dabei haben Wahrscheinlichkeiten, die den Annahmen zugrunde liegen, eine hohe Bedeutung. Da die Entscheidungsbaumanalyse eine mehrstufige Analyse ist, interessieren vor allem die Wahrscheinlichkeiten für das Eintreten eines Ereignisse B unter der

Voraussetzung, dass ein Ereignis A bereits eingetreten ist: Wie wahrscheinlich ist das Eintreten von B unter der Bedingung, dass A eingetreten ist. Diese Wahrscheinlichkeit wird als *bedingte Wahrscheinlichkeit* bezeichnet und ist mit dem Namen Thomas Bayes (Bayes-Formel) verbunden. Demnach kann man die Wahrscheinlichkeit nachfolgender (künftiger) Ereignisse (Ereignis B) auf der Grundlage bisheriger Erfahrungen (Ereignis A) abschätzen.[4]

Da die Entscheidungsbaumanalyse eine Entscheidungsfindung mittels Wahrscheinlichkeiten ist, kann sie als eine Form der Entscheidung unter Risiko angesehen werden. Entgegen der Bayes-Regel (siehe Abschnitt 3.1.2) werden die Ergebnisse über die Entscheidungsbaumanalyse ermittelt. Bei der Bayes-Regel wird eine Entscheidung auf Basis eines „fertigen" Entscheidungstableaus" getroffen.

### 3.2.2 Anwendung der Entscheidungsbaumanalyse

Zur Einführung in die Entscheidungsbaumanalysen soll folgendes Problem aus der betriebswirtschaftlichen Praxis dienen:

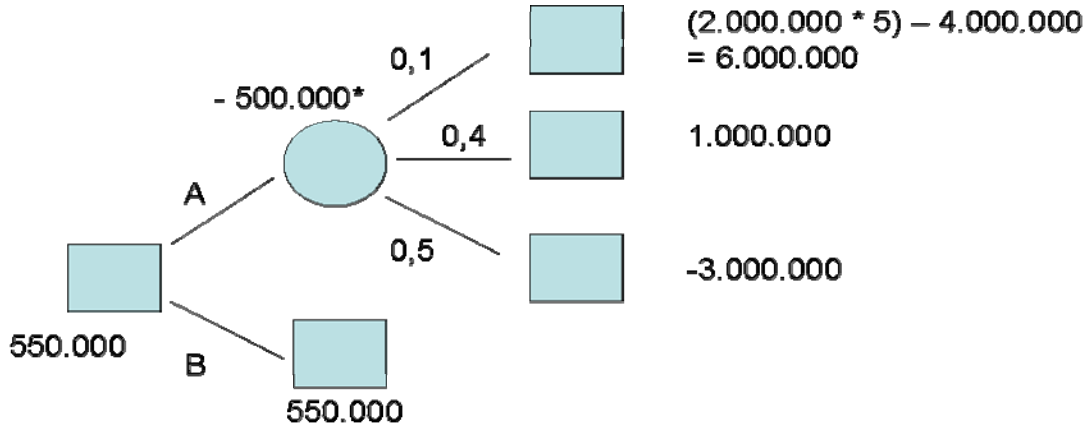
Ein Unternehmen plant die Einführung eines neuen Produkts und hält folgende drei Nachfrageniveaus $z_i$ mit entsprechenden Eintrittswahrscheinlichkeiten P ($z_i$) für möglich:

$z_1$ = 2.000.000 Stück, P ($z_1$) = 0,1
$z_2$ = 1.000.000 Stück, P ($z_2$) = 0,4
$z_3$ = 200.000 Stück, P ($z_3$) = 0,5

Der geplante Deckungsbeitrag je Einheit beträgt 5 GE und die fixen Kosten betragen 4 Millionen GE.
Alternativ zur Einführung eines neuen Produkts kann eine Investition getätigt werden, die einen sicheren Gewinn von 550.000 GE erbringt.
Soll sich das Unternehmen für das Projekt „Einführung eines neuen Produkts" (A) oder für die sichere Investition (B) entscheiden? Als Entscheidungsunterstützung soll die Entscheidungsbaumanalyse verwendet werden.

[4] Britische Wissenschaftlicher haben nachgewiesen, dass Tennisspieler die Bayes-Formel (unbewusst) anwenden, um die Bälle zurückzuschlagen. Um sich auf einen anfliegenden Ball einstellen zu können, müssen die Spieler die Richtung und die Geschwindigkeit abschätzen. Das Gehirn ist der Untersuchung zu folge in der Lage, auf Erfahrungen, die während des Spiels gemacht wurden, zurückzugreifen und Richtung und Geschwindigkeit des nächsten Balles mit einer hohen Wahrscheinlichkeit vorherzusagen. Siehe dazu den Bericht in der Berliner Zeitung vom 15. Januar 2004 beziehungsweise http://www.berlinonline.de/berliner-zeitung/archiv/.bin/dump.fcgi/2004/0115/wissenschaft/0057/index.html (Abruf: 14. August 2008).

Der Entscheidungsbaum sieht dann wie folgt aus:

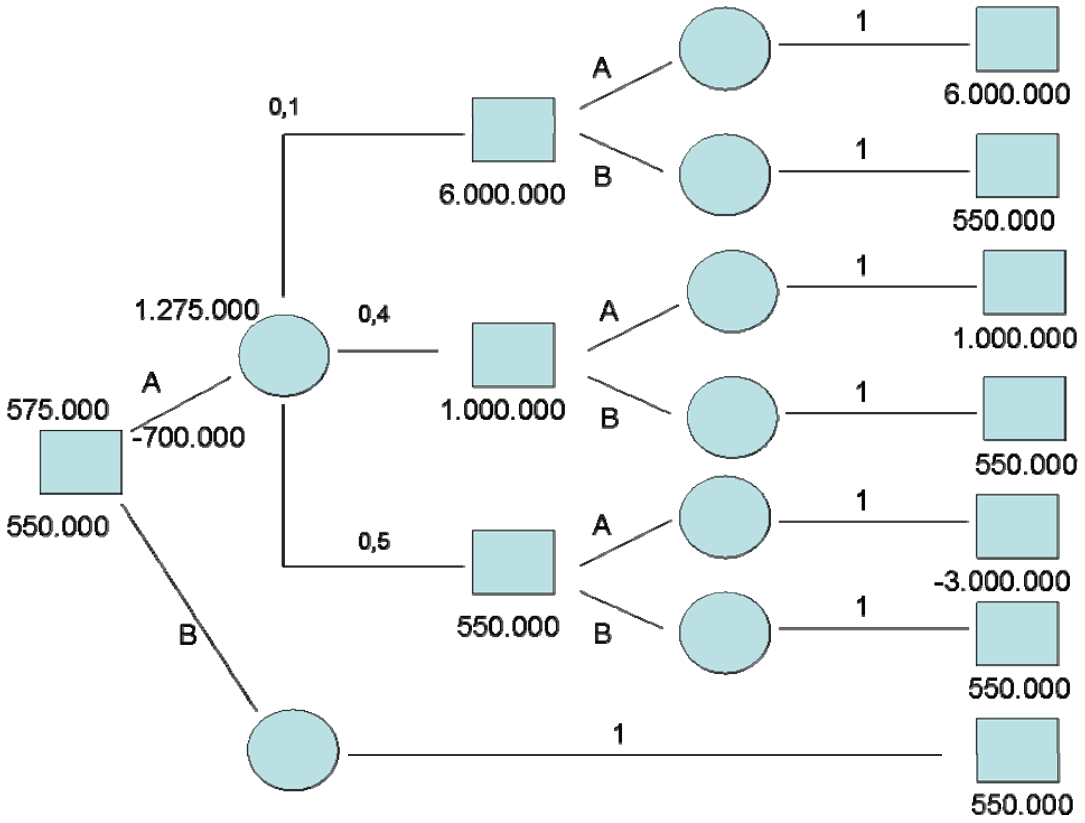



*** ergibt sich aus den Erwartungswerten abzüglich der Fixkosten**

Nach derzeitigem Informationsstand würde sich für die Investition mit dem sicheren Gewinn (B) entschieden werden.

Da bei betrieblichen Entscheidungen generell der Informationsstand zentral ist, können zusätzliche Informationen die Erwartungen „sicherer" machen. Informationen sind allerdings nicht kostenfrei verfügbar. Die Informationsökonomik spielt insbesondere in der Neuen Institutionenökonomik (etwa bei der Prinzipal-Agent-Theorie) eine zentrale Rolle. Kern dabei ist, dass es Informationsasymmetrien gibt und die Beschaffung von Informationen zur Entscheidungsunterstützung nicht unentgeltlich ist.

Im Beispiel soll nunmehr eine Marktanalyse erwogen werden, welche (nahezu) vollkommene Informationen über die Absatzchancen des Produktes liefert. Die Marktanalyse kostet allerdings 700.000 GE. Die Entscheidungsbaumanalyse muss nunmehr zum Ergebnis kommen, ob sich durch die zusätzlich beschaffte, nicht unentgeltliche Information eine andere Entscheidung (nämlich A statt B) ergibt.




Die Beschaffung vollkommener Informationen führt zu einer anderen Entscheidung. Die Einführung des neuen Produkts ist nunmehr der sicheren Investition vorzuziehen.

Die Bayes-Formel hat bis hierhin noch keine Rolle gespielt. Es gibt noch keine bedingten Wahrscheinlichkeiten. Die Erweiterung des Beispiels wird dies nun ändern.

Das Unternehmen habe die Möglichkeit, eine Testmarktuntersuchung für 50.000 GE von einem Fremdunternehmen durchführen zu lassen. In die Entscheidungsfindung des Unternehmens fließen die Resultate $y_1$ (günstiges Untersuchungsergebnis), $Y_2$ (durchschnittliches Untersuchungsergebnis) sowie $y_3$ (ungünstiges Untersuchungsergebnis) ein. Über die Verlässlichkeit der Testmarktuntersuchungen in Bezug auf die tatsächliche Marktsituation $z_i$ gibt es Erfahrungen in Form bedingter Wahrscheinlichkeiten:

| $P(y_i/z_i)$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $z_1$ | 0,6 | 0,3 | 0,1 |
| $z_2$ | 0,3 | 0,5 | 0,2 |
| $z_3$ | 0,1 | 0,2 | 0,7 |

Nochmals: $z_i$ repräsentieren die Wahrscheinlichkeiten für das Eintreten der verschiedenen Nachfrageniveaus. $y_i$ repräsentieren die Wahrscheinlichkeiten für die Testmarktergebnisse unter der Bedingung bestimmter Nachfrageniveaus $z_i$ ($P [y_i/z_i]$).

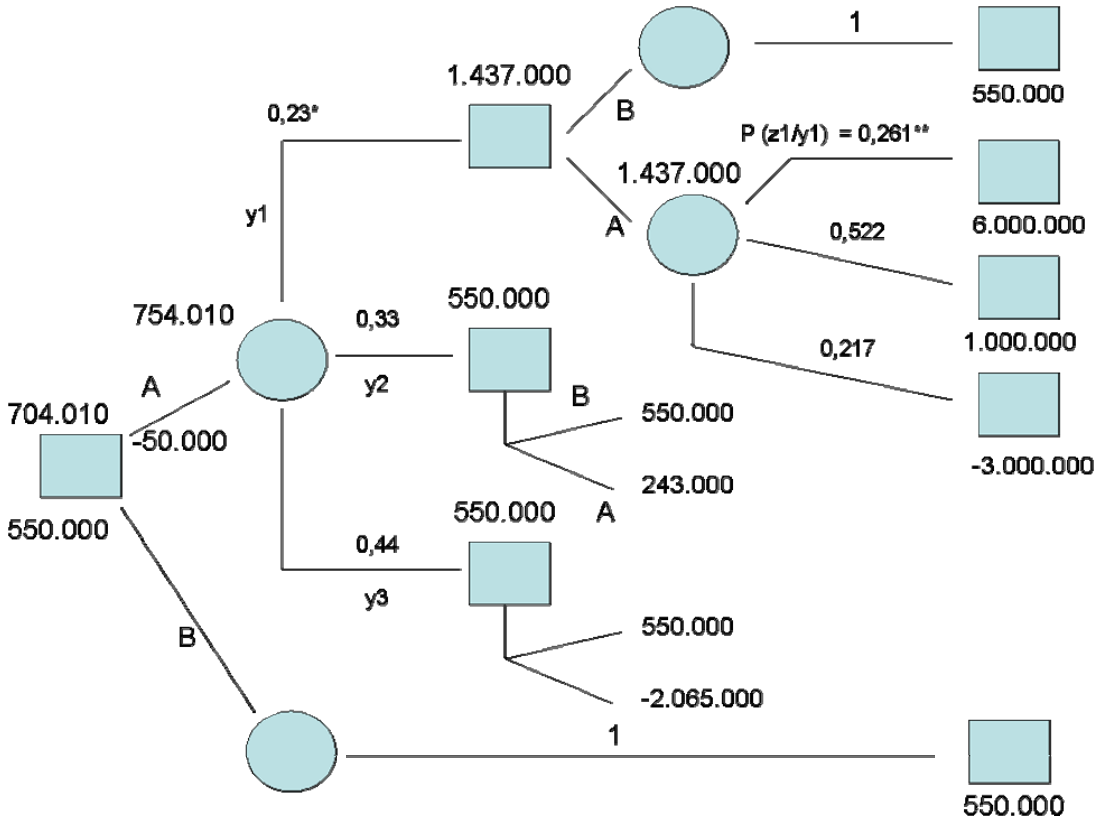
Die Wahrscheinlichkeiten für die bestimmten Ergebnisse der Testmarktanalyse ($P[y_i]$) ergeben sich (am Beispiel von $y_1$) aus

$$P(y_1) = P(y_1/z_1) * P(z_1) + P(y_1/z_2) * P(z_2) + P(y_1/z3) * P(z_3)$$

Zur Ermittlung der bedingten Wahrscheinlichkeiten wird die *Bayes-Formel* eingesetzt (am Beispiel von $y_1/z_1$):

$$P(z_1/y_1) = \frac{P(y_1 / z_1) * P(z_1)}{P(y_1)}$$

Der Entscheidungsbaum ergibt sich dann wie folgt:




* ergibt sich aus $Py_j = \sum_{1}^{3} P(y_j / z_i) * P(z_i)$, d.h. für $P(y_1)$: 0,6*0,1 + 0,3*0,4 + 0,1*0,5 = 0,23

** ergibt sich aus $P(z_1/y_1) = \frac{P(y_1 / z_1) * P(z_1)}{P(y_1)} = \frac{0,6 * 0,1}{0,23} = 0,261$

Die Entscheidungsbaumanalyse stellt eine strukturierte quantitative Bewertung von Entscheidungsalternativen dar und kann praktisch überall angewendet werden.

Das Problem der Entscheidungsbaumanalyse besteht auch hier darin, aus der Realität ein Modell abzuleiten, welches einerseits so nah wie nötig an der Realität, andererseits aber auch so überschaubar wie möglich ist.

## 3.3. Netzplantechnik

### 3.3.1 Grundlagen und Überblick

Die Netzplantechnik ist eine Form der Netzwerkanalyse. Bei Netzwerkanalysen geht es ganz generell um die Analyse von Beziehungsgeflechten innerhalb eines Netzwerks. Übertragen für Projekte geht es um die Analyse der Beziehungen zwischen Projektvorgängen.
Die Typologie der Netzwerkanalyse differenziert dabei zum Beispiel nach der Anzahl der beobachteten Beziehungen in

- uniplexe Netze: eine spezielle Beziehung im Netzwerk wird analysiert; im Projektumfeld könnte dies die spezielle Analyse eines (zentralen) Projektvorganges sein
- multiplexe Netze: mehrere Beziehungen im Netzwerk werden analysiert; im Projektumfeld könnte dies die Betrachtung eines Teilprojektes sein
- totale Netze: alle Beziehungen im Netzwerk werden analysiert; im Projektumfeld entspricht dies der Betrachtung des gesamten Projektes.

Auch die Maße zur Beschreibung von Netzwerken können wichtige Informationen für die Charakterisierung eines Projektes bereitstellen:
Die *Dichte* etwa misst den Verflechtungsgrad eines Netzes. Sie errechnet sich aus dem Verhältnis der empirisch vorhandenen Beziehungen zur Anzahl der grundsätzlich möglichen Beziehungen. Die Dichte kann Rückschlüsse auf die Komplexität und den Grad an Interdependenzen zwischen den Projektvorgängen zulassen.

Die *Zentralität* (Central Degree) misst die Stellung eines Knotens innerhalb eines Netzes und gibt damit Aufschluss über die Relevanz eines Knotens. Übertragen auf das Projektwesen kann es sich bei einem Vorgang mit hohem Central Degree um einen Vorgang handeln, auf den mehrere andere Vorgänge eine Anordnungsbeziehung haben. Klassicherweise ist dies bei Meilensteinen der Fall. Viele Vorgänge im Projekt müssen erfolgreich erledigt worden sein, damit ein Projektmeilenstein erreicht wird.

Komplexe Projekte, wie sie etwa Großprojekte der industriellen Fertigung darstellen können, erfordern die Erfüllung einer Vielzahl unterschiedlicher Arbeitspakete, die zeitlich und technisch interdependent sein können. Zur Strukturierung, Planung, Steuerung und Kontrolle solcher großer Projekte hat sich die Netzplantechnik als Hilfsmittel zur Optimierung bewährt. Die Netzplantechnik stellt ein Medium dar, welches es ermöglicht, etwa die Einhaltung der geplanten Termine oder die Determinierung kostengünstigsten Verkürzung der Projektdauer abzubilden.

Die Netzplantechnik stellt insgesamt eine Möglichkeit dar, Projekte (also betriebswirtschaftliche Anforderungen aus der realen Welt) in eine Modellform aus Knoten und Kanten zu überführen, um verschiedene projektspezifische Optimierungsnotwendigkeiten und damit -anforderungen zu lösen. Sie umfasst nach DIN 69900 alle Verfahren zur Analyse, Beschreibung, Planung, Steuerung, Überwachung von Abläufen auf der Grundlage der Graphentheorie und ist damit eine Form der Ablaufplanung, insbesondere im Umfeld betrieblicher Projekte.

Projekte sind zeitliche und von der Problemstellung begrenzte, vielfach einmalige Aufgaben, deren erfolgreiche Erfüllung vom Zusammenwirken verschiedener betriebsinterner wie betriebsexterner Stellen abhängt. Folglich sind deren Aktivitäten zu koordinieren. Der zeitlich a priori festgelegte Lebenszyklus eines Projektes macht die Notwendigkeit insbesondere zur Koordination und Optimierung der zeitlichen Restriktionen offenkundig: Viele Projekte haben, gegebenenfalls, einen Spielraum beim Projektbudget. Die wenigsten Projekte haben allerdings einen Spielraum bei der Erweiterung des Projektendes. Klassische Softwareprojekte sind etwa dadurch charakterisiert, dass der Entwicklungsschluss für alle Softwareprojekte, deren Ergebnisse in ein neues Software-Release einfließen, ein nicht änderbares Datum darstellt. Softwareprojekte, die nicht rechtzeitig fertig werden, können nicht Teil der Auslieferung des neuen Releases sein. Eine allgemeine Verschiebung eines Software-Releases und damit die Verschiebung der Bereitstellung neuer softwaretechnischer Funktionen für die darauf wartende Nachfrage verschlechtert die Marktchancen nachhaltig.

Ebenso stellt eine Kundenauftragsfertigung ein Projekt dar, in dem sich der Auftraggeber regelmäßig dazu verpflichten muss, zu einem (vertraglich) festgesetzten Termin zu liefern. Kann der Termin nicht eingehalten werden, drohen Konventionalstrafen.

Kern der Netzplantechnik ist, ein Projekt in einzelne in sich geschlossene Teilaufgaben (Arbeitspakete) aufzugliedern. Diese aufgegliederten Teilaufgaben werden in der Netzplantechnik als *Vorgänge* bezeichnet. Vorgänge, denen eine besondere Bedeutung beigemessen wird, werden als *Meilensteine* bezeichnet. Sie können zum Beispiel das Erreichen eines wichtigen Teilziels innerhalb des Gesamtprojektes symbolisieren (etwa das Erreichen eines fakturierungsfähigen Projektstandes/Produktionsstands).

Die funktionale Beziehung zwischen den Vorgängen wird in Form zeitlicher Abhängigkeiten im Netzplan abgebildet. Dieser Strukturierungsprozess dient der Erhöhung der Transparenz komplexer Projekte. Unter Umständen macht es diese Strukturierung überhaupt erst möglich, ein Projekt zu bewältigen. Auf jeden Fall ist die Schaffung von Transparenz die zentrale Voraussetzung für Optimierungen im Projekt, vor allem hinsichtlich der Zeitplanung, daneben aber auch hinsichtlich der Kapazitätsplanung, der Kostenplanung (hier ist die Optimierung ein Minimierungsproblem), der Steuerung sowie der Überwachung.

### 3.3.2 Theoretische Basis der Netzplantechnik: Die Graphentheorie

Wie jede Netzwerkanalyse handelt es sich bei der Netzplantechnik um die Analyse von Beziehungsgeflechten, also der Beziehungen zwischen einzelnen Arbeitspaketen eines Projektes in ihrer zeitlichen und/oder technologischen Bedingtheit. In einem Netzwerk und damit auch in einem Netzplan werden die Arbeitspakete als Knoten und die Beziehung mittels Kanten (Graphen) dargestellt. Im Sinne der mathematischen Graphentheorie stellt die Netzplantechnik ein endlicher, gerichteter und bewerteter Graph mit seiner Knotenmenge sowie seiner Menge gerichteter und bewerteter Kanten dar.
Die Knoten und Kanten des Netzplanes bilden im Projekt verschieden Größen dar, die das Projekt kennzeichnen. Dabei repräsentieren die Kanten („Strecken“) die Aktivitäten im Projekt (Arbeitsvorgänge) und die Knoten die Projekt-Ereignisse, also Anfang oder Ende dieser Aktivitäten.

Es lassen sich vier Anordnungsbeziehungen zwischen den Knoten unterscheiden:

- Ende-Anfang-Beziehung (Normalfolge)
- Anfang-Anfang-Beziehung (Anfangsfolge) Ausgangspunkt → zwei unabhängige Vorgänge können beginnen
- Ende-Ende-Beziehung (Endfolge): Zwei oder mehr Vorgänge müssen beendet sein, bevor das Projekt abgeschlossen werden kann
- Anfang-Ende-Beziehung (Sprungfolge)

Sämtliche Anordnungsbeziehungen stellen einen Mindestzeitabstand in Pfeilrichtung dar zwischen den jeweiligen Vorgangsereignissen. In der Netzplantechnik interessiert dabei stets der längste Weg im Graphen. Dieser wird als „kritischer Weg des Netzplanes" bezeichnet (Critical Path Method - CPM) und stellt den Mindestzeitbedarf für das Projekt dar. Die Existenz eines (endlichen) längsten Weges setzt dabei voraus, dass der Graph keine Zyklen enthält.

### 3.3.3 Anwendung der Netzplantechnik und Netzplanmethoden

Voraussetzung für die Anwendung der Netzplantechnik ist die Zerlegung des gesamten Projektes in seine kleinsten Ablaufelemente des Projektes, also in die Projektvorgänge. Diese Vorgänge einschließlich ihrer Anordnungsbeziehungen repräsentieren schließlich die Planungsobjekte.

Mit der Projektanalyse ist bereits ein wesentlicher Teil der Netzplanerstellung erarbeitet. In der Praxis erweist sich diese Arbeit allerdings regelmäßig als komplexe und schwierige Aufgabe. Ungenauigkeiten, Unvollständigkeiten und Fehler bei der Projektzerlegung können zu folgeschweren Konsequenzen bis hin zum Misserfolg des Projekts führen. Klassisch wird hierzu immer wieder auf die so genannte „80-20-Regel" verwiesen. Mit der Strukturanalyse, also der vollständigen (und korrekten!) Ermittlung und Erfassung der Vorgänge einschließlich ihrer Anordnungsbeziehungen, werden 80% des Erfolgs des Einsatzes der Netzplantechnik erzielt und damit ein erheblicher Beitrag zum Projekterfolg insgesamt geleistet.

Neben der Projektplanung auf Basis der Netzplantechnik darf nicht vergessen werden, dass die Projektsteuerung während der Projektrealisierung sowie die Projektkontrolle im gesamten Lebenszyklus eines Projektes eine zentrale Rolle hinsichtlich des Projekterfolgs spielen. [5]

Die systematische Strukturierung eines Projektes für die Planung, Realisierung und Kontrolle erfolgt im *Projektstrukturplan*. Dieser stellt nach DIN 69901 die Darstellung der Projektstruktur dar und kann nach dem Projektablauf eben als Netzplan aufgestellt werden. Andere Formen der Projektstrukturierung sind vor allem jene nach dem Aufbau (Aufbaustruktur) und nach Grundbedingungen (Grundstruktur, Wahlstruktur). In jedem Fall soll mit dem Projektstrukturplan die Vollständigkeit der Erfassung und Beschreibung sichergestellt werden. Dabei ist vor allem auch die Frage nach der Gliederungstiefe, also nach dem Grad der Granularität der Projektanalyse von Relevanz. Fragen nach der Aggregation und Komplexitätsreduktion sind bei der Projektzerlegung sinnvoll zu beantworten.

Die Gliederungstiefe sollte im gesamten Projektstrukturplan gleich sein. Die Arbeitspakete sollten demnach nur auf der untersten Projektstrukturebene erscheinen und nicht weiter aufgegliedert werden. Im aus dem Projektstrukturplan abzuleitenden Netzplan sind dann diese

[5] Der Projekterfolg ist allerdings von verschiedensten Gegebenheiten abhängig.

Arbeitspakete in mehrere Vorgänge zu zerlegen. In diesem Sinne ist ein Arbeitspaket sogar als Teilnetz zu verstehen.

Bei Erstellung des Netzplans können in einer einfachen Tabelle für jeden Vorgang die Zeitdauern angegeben werden. Es hat sich der besseren Übersichtlichkeit halber bewährt, zu jedem Vorgang die unmittelbar vor- und nachgeordneten Vorgänge anzugeben. Ausgehend aus dieser Liste ist es dann möglich, den Netzplan zu zeichnen. Der Netzplan stellt letztlich nichts anderes dar als die graphische Wiedergabe der Zusammenhänge des betrachteten Projekts. Zur Auswertung des Netzplanes können dann verschiedene Netzplanmethoden angewendet werden. Weil bei Projekten die Zeitoptimierung im Mittelpunkt steht, interessiert zunächst, welche Mindestdauer für die Projektdurchführung notwendig ist. Dazu müssen alle potentiell möglichen Wege, dies sind lückenlose Abfolgen von Pfeilen, vom Ausgangsknoten bis hin zu Endknoten bestimmt und die Gesamtzeitdauern aller dazugehörigen Tätigkeiten ermittelt werden. Die Methode des so genannten „Kritischen Pfads" (Critical Path Method - CPM) ermittelt dann jenen Weg im Netzplan aus, der die größte zeitliche Dauer und damit den geringsten zeitlichen Puffer aufweist. Die diesen Weg abbildenden Projektvorgänge sind für die Projektdurchführung und damit für den Projekterfolg von entscheidender Relevanz: Verzögerungen im Beginn sowie in der Durchführung dieser Vorgänge führt zu einer Verlängerung der Gesamtprojektdauer.

Folgendes Beispiel eines Projektes soll die Netzplantechnik am Beispiel der Anwendung der Critical Path Method[6] verdeutlichen:

Das Projekt besteht aus sechs Projektvorgängen: PV1 bis PV6. Die Anordnungsbeziehungen im Projekt ergeben sich wie folgt:

- PV1 dient der Projektvorbereitung. Sie erfordert 10 Zeiteinheiten (ZE). PV2 folgt auf PV1.
- PV3 kann frühestens fünf ZE nach Beginn von PV2 anfangen, muss andererseits aber spätestens 10 ZE nach Beginn von PV5 angefangen werden
- PV4 kann frühestens beginnen, wenn PV3 zu ¾ abgearbeitet ist und spätestens 10 ZE nach dem Anfang von PV6
- Die Anfänge von PV5 und PV6 dürfen höchstens 10 ZE auseinander liegen.
- Spätestens 15 ZE nachdem PV5 zur Hälfte abgearbeitet wurde, muss PV2 beginnen.

Die jeweiligen Projektvorgangsdauern seien wie folgt vorgegeben:

| Vorgang | PV1 | PV2 | PV3 | PV4 | PV5 | PV6 |
|---|---|---|---|---|---|---|
| Dauer in ZE | 10 | 10 | 20 | 20 | 10 | 5 |

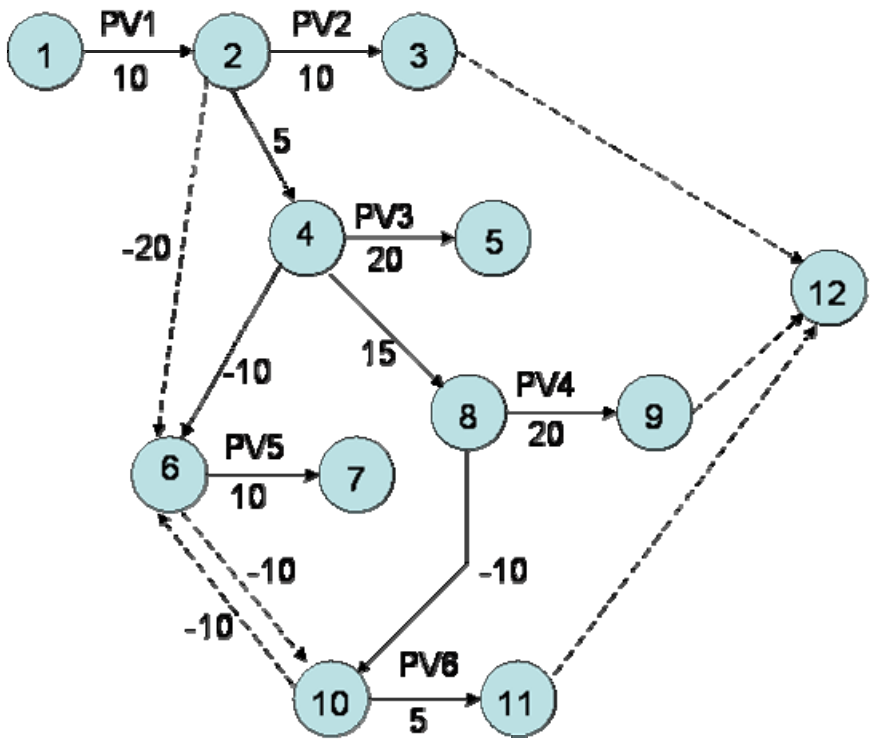
Bei der CPM handelt es sich um einen so genannten Vorgangspfeil-Netzplan. Die Vorgänge werden durch Pfeile dargestellt und jeder Pfeil umgekehrt als Vorgang aufgefasst. Die Knoten eines CPM-Netzplans werden als Ereignisse bezeichnet.

[6] Als weitere Netzplanmethoden wären zum Beispiel zu nennen: Metra Potential Method, Hamburger Methode der Netzplantechnik oder Programm Evaluation and Review Technique.

Die Erstellung eines CPM-Netzplans erfolgt anhand verschiedener Konventionen:

- Münden mehrere Pfeile in einen Knoten, so bedeutet dies, dass das korrespondierende Ereignis dann eintritt, wenn alle Vorgänge, die den einmündenden Pfeilen entsprechen, beendet sind. Ein Knoten, in den kein Pfeil einmündet, wird als Startknoten und damit als Startereignis interpretiert. Ein Startereignis gilt dabei stets als eingetreten.
- Ein Vorgang kann erst dann anfangen, wenn das Ereignis, welches Ausgangspunkt des Vorgangspfeils ist, eingetreten ist.
- Es ist, wie im Beispiel, notwendig, Vorgangsüberlappungen („PV4 kann erst dann beginnen, wenn PV3 zu ¾ abgearbeitet ist"), Kopplungen („die Anfänge von PV5 und PV6 dürfen höchstens 10 ZE auseinander liegen") sowie Wartezeiten im CPM-Netzplan abzubilden. Dies erfolgt mittels so genannter *Scheinvorgänge*. Ein solcher Scheinvorgang kann eine Vorgangsdauer, die Null, positiv oder negativ sein kann. Ein nicht bewerteter Vorgang innerhalb eines CPM-Vorgangs hat die Dauer „Null".

Der CPM-Netzplan aus dem Beispiel ergibt sich dann wie folgt:




Nach der Abbildung des Netzplans können die Netzplananalysen beginnen. Im Rahmen der vor allem interessierenden Terminplanung werden den Knoten des Netzplans frühestmögliche Zeitpunkte bezogen auf den Projektbeginn zum Zeitpunkt „Null" zugeordnet sowie spätestmögliche ohne Projektverzögerungen. Der Algorithmus zur Terminplanung ist grundsätzlich unabhängig davon, welche Netzplanmethode gewählt wird.

Zur Bestimmung des frühestmöglichen Anfangszeitpunkts (FAZP) setzt man PV1 $FAZP_{PV1} = 0$ und bestimmt vom Startknoten ausgehend, also von PV1 aus, den längsten Weg zu jedem Knoten des Netzplans. Unter einem Weg innerhalb eines Graphen im Sinne der Graphentheorie versteht man eine durch die Pfeilrichtung determinierten Folge, welche aus einer Sequenz

von Knoten und Pfeile besteht. Die Weglänge ergibt sich als Summe der einzelnen Pfeilbewertungen des Weges. Der längste Weg vom Projektstart zu einem Knoten determiniert den frühesten Zeitpunkt zu dem das korrespondierende Ereignis stattfinden kann. Infolgedessen werden, ausgehend vom Projektanfang PV1 sukzessive die längsten Wege ermittelt (so genannte Vorwärtsterminierung): $FAZ_{PV2}$ = 10, $FAZ_{PV3}$ = 15, $FAZ_{PV4}$= 30. Zum Knoten D führen mehrere Wege: PV2 → PV5 = -10, PV3 → PV5 = 5 und über PV6 → PV5 = 10. $FAZ_{PV5}$ ist demnach der längste Weg, also 10.

Der längste Weg von PV1 nach PV6 verläuft über den Knoten PV4 und hat die Länge $FAZ_{PV6}$ = 20.

Schließlich kann nun der längste Weg von PV1 bis zum Projektende berechnet werden. Dieser führt über den Knoten PV4 und hat die Länge $FAZ_{Projektende}$ = 50 ($FAZ_{PV4}$ + $FAZ_{PV6}$ = 30 + 20 = 50). Das Projekt kann also frühestens 50 ZE nach dem Projektstart beendet werden.

Zusätzlich zur Frage, wann ein Projektvorgang frühestens anfangen kann, stellt sich auch die Frage nach dem spätesten Anfangszeitpunkt (SAZP), ohne dass dadurch das Projektende gefährdet wird. Die Frage wird dadurch beantwortet, dass man ausgehend vom Projektende schrittweise geeignete Vorgängerknoten sucht, um den jeweils längsten Weg zum Projektende zu bestimmen (Rückwärtsterminierung). Gesucht ist der längste Weg eines beliebigen Knotens im Netzplan zum Projektende. Er gibt an, wie viele ZE nach dem Anfang eines Knotens bis zum Projektende verstreichen müssen. $SAZP_{Projektende}$ = $FAZP_{Projektende}$ = 50. Vom Knoten PV4 aus geht ein Pfeil zum Projektende mit der Weglänge 20. Des weiteren gibt es vom Knoten PV4 aus einen Weg über PV6 mit einer Weglänge von -5 sowie einen längsten Weg über PV5 mit -10 Weglänge. Der längste Weg von PV4 zum Projektende hat also die Länge 20. Daher ist $SAZP_{PV4}$ festgesetzt mit 50 -20 = 30 = $FAZP_{PV4}$.

Der frühestmögliche Anfangstermin sowie der spätest zulässige Endtermin sind für die Berechnung der Pufferzeit vonnöten.

Nun sind alle Voraussetzungen geschaffen, den maximalen Dispositionsspielraum im Projekt für den jeweiligen Vorgangsbeginn. Dieser maximale Dispositionsspielraum wird als „gesamte Pufferzeit" (GPZ) bezeichnet. Sie ist das Ergebnis aus SAZP – FAZP. Vorgänge mit einer gesamten Pufferzeit von Null sind nicht disponibel und daher kritisch.

Eine zusammenfassende Übersicht über FAZP, SAZP und GPZ ergibt folgendes Bild[7]:

| Vorgang | FAZP | SAZP | GPZ |
|---|---|---|---|
| PV1 | 0 | 0 | 0 |
| PV2 | 10 | 10 | 0 |
| PV3 | 15 | 15 | 0 |
| PV4 | 30 | 30 | 0 |
| PV5 | 10 | 40 | 30 |
| PV6 | 20 | 45 | 25 |
| Projektende | 50 | 50 | 0 |

Es lässt sich nunmehr ableiten, dass eine Verzögerung von Vorgang PV6 trotz seines Puffers von 25 Zeiteinheiten den frühesten Anfangszeitpunkt von PV5 ($FAZ_{PV5}$) um die gleiche ZE ver-

[7] Der besseren grundsätzlichen Übersicht halber werden weitere Arten von Pufferzeiten nicht aufgeführt.

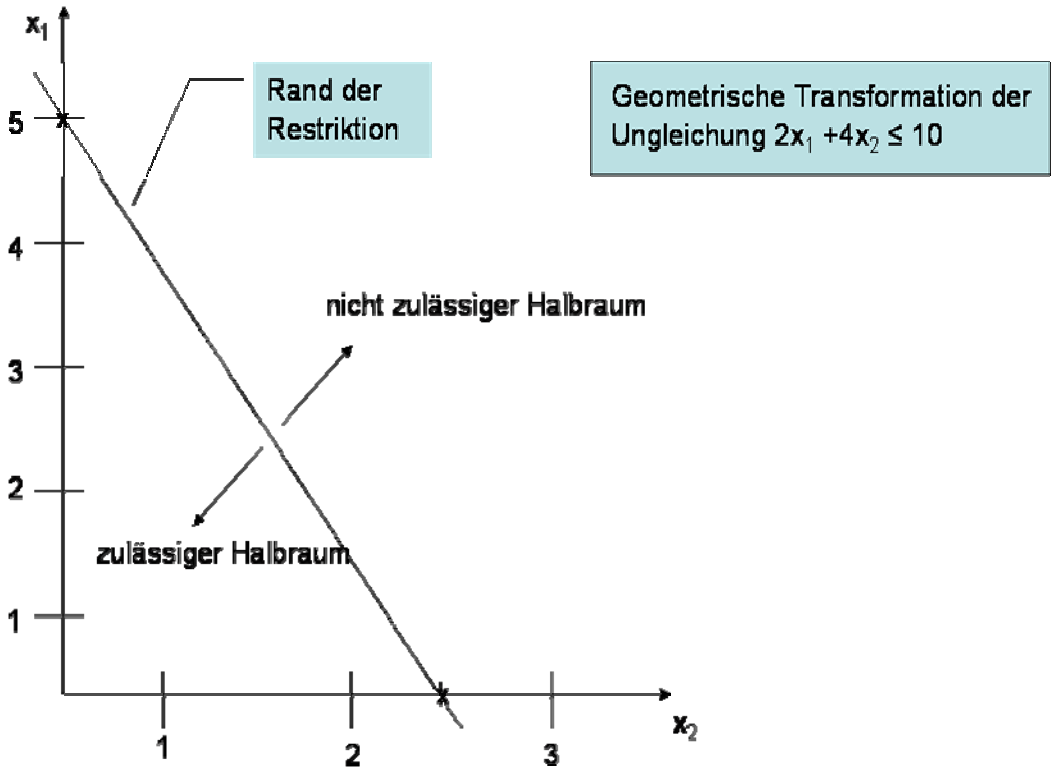
schieben würde. Die freie Pufferzeit ist dann jene Zeitspanne, um die der Vorgang gegenüber seiner frühesten Lage verschoben werden könnte, ohne die früheste Lage anderer Vorgänge zu beeinflussen.

## 3.4. Lineare Optimierung

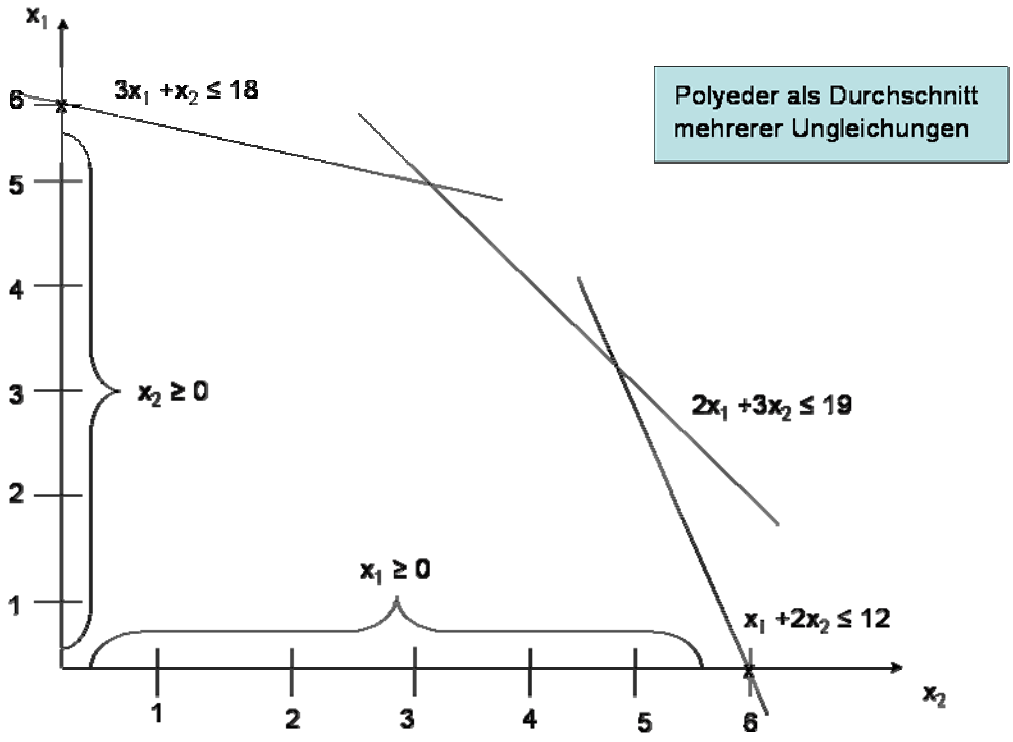
### 3.4.1 Grundlagen und Überblick

Die lineare Optimierung, auch lineare Programmierung genannt (hier „Programmierung" als synonym für Planung), umfasst Methoden und Verfahren zur Lösung der Optimierungsaufgabe, eine lineare Zielfunktion zu maximieren oder zu minimieren (also zu optimieren) unter Berücksichtigung linearer Nebenbedingungen (also Restriktionen). Die in der Zielfunktion zu optimierenden Variablen müssen dem Kriterium der Nichtnegativitätsbedingung genügen. In der Regel stellen nämlich die Variablen Outputmengen dar, die nur Sinn machen, wenn sie ≥ Null sind.

Die Nebenbedingungen können der Art ≤, ≥ und/oder = sein. Die *geometrische Analyse* als Äquivalent einer Ungleichung teilt einen n-dimensionalen Raum in zwei Halbräume auf. Dabei wird einer der beiden Halbräume durch < beziehungsweise > zugelassen, der andere verboten. Die dem Gleichheitszeichen entsprechende Ebene bildet dann den Rand des zulässigen Halbraums:



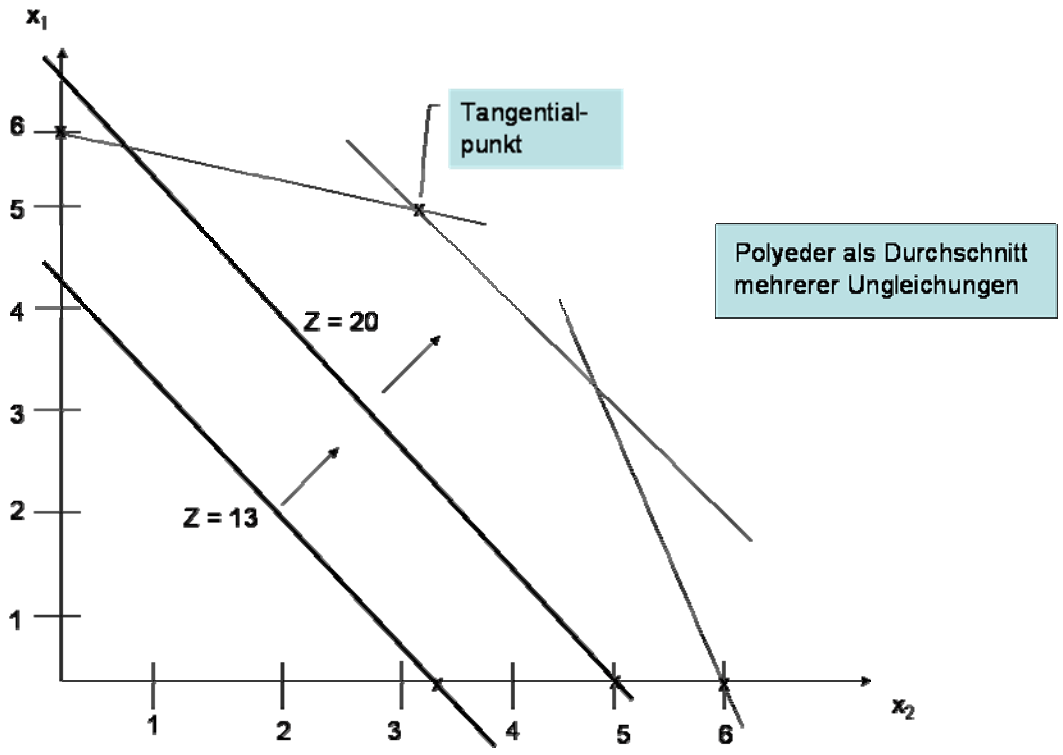
Sind mehrere Restriktionen (Nebenbedingungen) beziehungsweise Nichtnegativitätsbedingungen gegeben, so reduziert sich das Gebiet der zulässigen Lösungen auf den Durchschnitt der zulässigen Halbräume. Der Bereich, der alle Restriktionen erfüllt, wird als Polyeder bezeichnet. Die nachfolgende Abbildungen stellen graphisch den Lösungsraum dar für die Zielfunktion $Z = 3x_1 + 4x_2 \rightarrow$ max!. Der Lösungsraum wird durch die angegebenen linearen Nebenbedingungen determiniert:




Die Zielfunktion kann, bei alternativen Werten für Z, in Form von Isoquanten der Zielfunktion abgebildet werden. Die Isoquanten sind zueinander parallel. Der optimale (das heißt maximale oder minimale) Zielwert ist an jenem Punkt des Polyeder ermittelt, an dem eine der Isoquanten den Polyeder gerade noch tangiert. Man verschiebt als die Isoquante so lange hin zu den Kanten des Polyeders, bis jener Punkt erreicht wird, der den maximalen Zielfunktionswert generiert. Dieser Tangentialpunkt ist die gesuchte Optimallösung. Daraus folgt des Weiteren, dass die Optimallösung mindestens an einer Ecke des den Lösungsraum abbildenden Polyeders eintritt. Eine Ecke ist dann optimal, wenn alle zu ihr entlang von Kanten benachbarten Ecken des Polyeders schlechtere Zielwerte haben. Der Schnittpunkt des Polyeders ist dabei stets der Schnittpunkt von zwei der Nebenbedingungen, die den Polyeder insgesamt formen. Dies bedeutet, dass die anderen Nebenbedingungen irrelevant sind für die Ermittlung des Optimums:

Die rechentechnische Ermittlung der optimalen Ecke im Polyeder erfolgt mittels der Simplex-Methode. Diese stellt einer Erweiterung der Verfahren zur Lösung linearer Gleichungssysteme dar (wie sie etwa auch im Gleichungsverfahren bei der Ermittlung der Leistungspreise einander gegenseitig Leistungen austauschenden Kostenstellen zum Ausdruck kommt).

Die rechentechnische Ermittlung der Optimallösung setzt dabei eine vorangegangene graphische Abbildung des Tangentialpunkts voraus. Nur so ist es möglich, die relevanten Nebenbedingungen zu detektieren und in den Simplex-Ansatz einfließen zu lassen.




Für die rechentechnische Ermittlung der Optimallösung wird die Standardform des Linearen Programmierungs (LP)-Modells als System von Ungleichungen wie folgt abgebildet:

| | |
|---|---|
| $Z = c^T * \rightarrow$ ‚optimiere!' | Zielfunktion in Matrixschreibweise[8]<br>mit $c_i$ = Zielfunktionskoeffizienten |
| $A * x \leq b$ | Nebenbedingung(en) |
| $x \geq 0$ | Nichtnegativitätsbedingung |

Die Normalform des LP-Modells ergibt sich in Form eines Systems von Gleichungen wie folgt:

| | |
|---|---|
| $Z = c^T * \rightarrow$ ‚optimiere!' | Zielfunktion |
| $A * x + S = b$ | Nebenbedingung(en) mit S = Schlupfvariable[9] |
| $x, S \geq 0$ | Nichtnegativitätsbedingung |

[8] In Gleichungsform ergibt sich analog: $Z = c_1*x_1 + c_2*x_2 + \ldots + c_n*x_n$, mit
$a_{1,1}*x_1 + a_{1,2}*x_2 + \ldots a_{1,n}*x_n \leq b$ und
$a_{m,1}*x_1 + a_{m,2}*x_2 + \ldots a_{m,n}*x_n \leq b_m$.
[9] Siehe dazu Abschnitt 3.3.3.

In Tabellenform transferiert kann ein LP-Modell wie folgt dargestellt werden:

| | $x_1$ | $x_2$ | $x_3$ | $x_i$ | Art der NB[10] | b |
|---|---|---|---|---|---|---|
| Nebenbedingung$_1$ | $a_{1,1}$ | $a_{1,2}$ | … | $a_{1,i}$ | | $b_1$ |
| $NB_2$ | $a_{2,1}$ | … | … | … | | $b_2$ |
| $NB_3$ | … | … | … | … | | $b_3$ |
| $NB_n$ | $a_{n,1}$ | … | … | $a_{n,i}$ | | $b_n$ |
| Zielfunktion | $c_1$ | $c_2$ | … | $c_n$ | | |

mit $a_{n,i}$ = Koeffizienten der Nebenbedingungen

### 3.4.2 Anwendung der linearen Optimierung

Klassische betriebswirtschaftliche Problemfelder können Anwendungsgebiet der Methode der linearen Optimierung sein, vor allem die betriebliche Fertigung. Insbesondere bei Mehrproduktunternehmen kann es Probleme beim Produktionsprogramm geben. Eine Anlage kann etwa mehrere Produkte herstellen. Wie legt man nun fest, wie man die Maschine einsetzt? Die Maschine hat eine festgelegte Kapazität (Maschinenstunden, Menge produzierter Leistungseinheiten)?

Die Anwendung der Simplex-Methode soll anhand des nachfolgenden Beispiels veranschaulicht werden. Es wird wieder der Grundgedanke des Operations Research insgesamt deutlich: Ein betriebswirtschaftliches Problem aus der realen Welt wird in ein Modell von Ungleichungen übersetzt, welches graphisch und rechentechnisch gelöst werden kann und als Ergebnis die Optimallösung als Entscheidungsgrundlage bereithält.

Folgendes Beispiel zeigt die Umwandlung eines betriebswirtschaftlichen Problems in eine Zahlenwelt und die Lösung des Problems:

*Ein Unternehmen fertigt die Produkte X und Y, welche die Deckungsbeiträge 300 beziehungsweise 500 Geldeinheiten/Stück haben. Mit den Deckungsbeiträgen sind die fixen Kosten in Höhe von 36.000 Geldeinheiten zu decken.*

*Die Ausbringungsmengen sind durch die Kapazitäten der Fertigungsanlagen determiniert. Produkt X durchläuft die Maschinen 1 und 2, die es jeweils eine Stunde je Mengeneinheit belegt. Für die Produktion von Produkt Y werden je Mengeneinheit zwei Stunden auf Maschine 1, eine Stunde auf Maschine 2 und drei Stunden auf Maschine 3 benötigt. Die maximalen Kapazitäten für die Maschinen 1, 2 und drei betragen 170, 150 beziehungsweise 180 Stunden.*

Das zu lösende Optimierungsproblem ist die Suche nach derjenigen monatlich zu produzierenden Mengen der Produkte X und Y, für die sich unter Beachtung der Kapazitätsrestriktionen ein maximaler Gewinn ergibt unter der Annahme, dass die direkten Kosten und Erlöse je Mengeneinheit mengenunabhängig sind. Darüber hinaus gibt es keine zu berücksichtigende Absatzgrenzen.

[10] ≤ beziehungsweise ≥.

Lösungsschritte:

1. Die Modellösung beginnt mit der Aufstellung der Zielfunktion. Sie lautet:

$$300x + 500y - 36.000 \rightarrow \max!$$

mit x = Produkt 1 und y = Produkt 2

2. Zur Zielfunktion sind die Nebenbedingungen aus dem realen betriebswirtschaftlichen Problem zu formulieren. Sie betreffen die Kapazitätsrestriktionen der Maschinen und lauten:
Für Maschine 1: $x + 2y \leq 170$
Für Maschine 2: $x + y \leq 150$
Für Maschine 3: $3y \leq 180$

Ferner gilt die Nichtnegativitätsbedingung, da es sich bei x und y um Produkte beziehungsweise Produktmengen handelt, die zu optimieren sind:
$x, y \geq 0$

Annahme: Nebenbedingungen $x + 2y \leq 170$ und $x + y \leq 150$ sind relevant.

Daraus folgt:
$x = 170 - 2y$

$x = 130$
$y = 20$

Die monatlich zu erzeugenden Mengen der Produkte X und Y lauten 130 beziehungsweise 20. Damit wird ein maximaler Netto-Erlös in Höhe von 13.000 Geldeinheiten erreicht (Einsetzen der Werte in die Zielfunktion).

Optimierungsrechnungen versuchen weiterhin, die Auswirkung von Änderungen in den Annahmen auf das Optimalergebnis mit Hilfe von Sensitivitätsanalysen zu analysieren. Hier können zum Beispiel die Auswirkungen angenommener Preisänderungen berechnet werden. Allgemein könnte man sagen, die Sensitivitätsanalyse kann Auskunft über die Reagibilität („Empfindlichkeit") der Optimallösung gegenüber Parameteränderungen geben.

### 3.4.3 Lösung einer linearen Optimierung mittels der Simplexmethode

Die Simplexmethode, eine Anwendung der Lagrange'schen Multiplikation, wandelt die Ungleichungen in Gleichungen durch Einführen zusätzlicher Variablen um. Diese zusätzlichen Variablen werden als *Schlupfvariablen* bezeichnet. Sie sorgen -formal-, dafür, dass die linke Seite der Ungleichung genauso groß wird wie die rechte Seite. Es werden genauso viele Schlupfvariable eingeführt wie es Nebenbedingungen gibt. Die Schlupfvariablen selbst sind keine entscheidungsrelevanten Variablen. Sie verändern daher den Zielwert nicht oder haben, anders ausgedrückt, in der Zielfunktion den Wert Null. Die entscheidungsrelevanten Variablen werden Strukturvariablen genannt.

Jede Ungleichung der Form „≤" wird in eine Gleichung transferiert, indem eine Schlupfvariable addiert wird (die linke Seite der Gleichung ist potentiell kleiner als die rechte Seite der Gleichung). Jede Ungleichung der Form ≥ wird in eine Gleichung transferiert, indem die Schlupfvariable subtrahiert wird (die linke Seite der Gleichung ist potentiell größer als die rechte Seite der Gleichung).

Als Beispiel zur Darstellung des Simplexverfahrens soll folgende Zielfunktion dienen:

$12x_1 + 15x_2 \rightarrow$ max! mit den Nebenbedingungen
$8x_1 + 6x_2 \leq 120$ und
$4x_1 + 10x_2 \leq 100$
$x_1, x_2 \geq 0$

Die Ausgangsdaten werden in ein Tableau eingestellt und gleichzeitig unter Verwendung von Schlupfvariablen zu Gleichungen transformiert:

| | $X_1$ | $X_2$ | $Y_1$ | $Y_2$ | b |
|---|---|---|---|---|---|
| $Y_1$ | 8 | 6 | 1 | 0 | 120 |
| $Y_2$ | 4 | 10 | 0 | 1 | 100 |
| Z | -12 | -15 | 0 | 0 | 0 |

Die Lösung des Tableaus erfolgt mittels des Simplexalgorithmus durch die Anwendung des Gauß-Yordan-Verfahrens.

- Im ersten Schritt dieses Verfahrens wird jene Spalte des Tableaus mit dem höchsten negativen Wert in der Zielfunktionszeile bestimmt. Diese Spalte wird als Pivotspalte bezeichnet.
- Im zweiten Schritt wird jene Zeile bestimmt, bei der der Quotient aus dem Koeffizienten der Ergebnisspalte (b) und dem Koeffizienten in der Pivotspalte den geringsten positiven Wert annimmt. Diese Zeile wird als Pivotzeile bezeichnet.
- Das so gefundene Pivotelement (Basisvariable) wird dann aus der Lösung entfernt.
- Zur Ermittlung der Basislösung werden Zeilenoperationen ausgeführt
    - Teile das Pivotelement durch sich selbst und die anderen Zeilenwerte durch das Pivotelement ($\rightarrow$ das Pivotelement nimmt den Wert 1 an)
    - Subtrahiere von den restlichen Elementen der Pivotspalte sich selbst ($\rightarrow$ das Element [M] nimmt den Wert 0 an)
    - Bestimme die restlichen Elemente nach dem Schema:

$$Wert_{neu} = Wert_{alt} - (M * \frac{"Tauschwert"}{Pivotelement})$$

Der „Tauschwert“ ist der Wert des jeweiligen Elementes der Pivotzeile.

Pivotspalte, Pivotzeile und Pivotelement sind nachfolgend hervorgehoben:

| | $X_1$ | $X_2$ | $Y_1$ | $Y_2$ | b |
|---|---|---|---|---|---|
| $Y_1$ | 8 | 6 | 1 | 0 | 120 |
| $Y_2$ | 4 | **10** | 0 | 1 | 100 |
| Z | -12 | -15 | 0 | 0 | |

Die Durchführung des Gauß-Yordan-Verfahrens führt zu einem Tableau 1:

| | $X_1$ | $X_2$ | $Y_1$ | $Y_2$ | b |
|---|---|---|---|---|---|
| $Y_1$ | $\frac{56}{10}$ [11] | **0** | 1 | $-\frac{6}{10}$ [12] | 60 |
| $Y_2$ | $\frac{4}{10}$ | 1 | 0 | $\frac{1}{10}$ | 10 |
| Z | -6[13] | **0** | 0 | $\frac{3}{2}$ [14] | |

Tableau 1 stellt noch nicht die Optimallösung dar. Dazu müssten alle Zielwerte positiv sein. Es ist daher ein Tableau 2 aufbauend auf Tableau 1 zu erstellen, in dem die Schritte zur Ableitung von Tableau 1 (Bestimmung von Pivotspalte, Pivotzeile, Pivotelement) wiederholt werden. Für Tableau 2 ergibt sich:

| | $X_1$ | $X_2$ | $Y_1$ | $Y_2$ | B |
|---|---|---|---|---|---|
| $Y_1$ | 1 | 0 | $\frac{10}{56}$ | $\frac{-6}{56}$ | $\frac{600}{56}$ |
| $Y_2$ | 0 | 1 | $\frac{-4}{56}$ | $\frac{8}{56}$ | $\frac{40}{7}$ |
| Z | 0 | 0 | $\frac{60}{56}$ | $\frac{60}{70}$ | |

Alle Zielwerte sind positiv: Tableau 2 ist das Optimaltableau. In den Zeilen werden $y_1$ beziehungsweise $y_2$ mit $x_1$ und $x_2$ ersetzt:

$x_1 = \frac{600}{56}$ und $x_2 = \frac{40}{7}$

In die Zielfunktion eingesetzt ergibt sich ein Zielfunktionswert von 214,29.

---

[11] $8 - (6 * \frac{4}{10})$

[12] $0 - (6 * \frac{1}{10})$

[13] $-12 - (-15 * \frac{4}{10})$

[14] $0 - (-15 * \frac{1}{10})$

### 3.4.4 Schattenpreise (Grenzproduktivität, Opportunitätskosten)

Gegeben sei folgendes bereits in Form der linearen Programmierung transformiertes betriebswirtschaftliches Maximierungsproblem:

G = 7x + 10y mit den Nebenbedingungen
3x + 2y ≤ 36,
2x + 4y ≤ 40,
x ≤ 10,
x, y ≥ 0

Dieses primale Problem kann als duales Problem formuliert werden. Generell gilt dabei, dass ein primales Minimierungsproblem in ein duales Maximierungsproblem umgewandelt werden kann und primales Maximierungsproblem in ein duales Minimierungsproblem. Das duale Problem ist die logische Umkehrung des ursprünglichen (= primalen) Problems. Wenn das primale Problem etwa die Maximierung des Gewinns bei gegebenen Kapazitätsbeschränkungen ist, so das duale Problem und damit die Umkehrung des Ursprungproblems die Minimierung der Kosten bei gegebenen Ziel.

Die Variablen des dualen Problems (Dualvariablen) haben dann eine ökonomisch interessante Bedeutung: Sie sind die Schattenpreise.

Das primale Problem lässt sich als Modell und in Tabellenform wie folgt darstellen:

$$Z = \sum_{\substack{n=1 \\ i=1}}^{m} c_n * x_i \rightarrow \text{max!}$$

mit

$$\sum_{\substack{n=1 \\ i=1}}^{m} a_{i,n} * x_i \leq b_n$$

mit $x_i \geq 0$

| $x_1$ | $x_2$ | $x_3$ | $x_i$ | Art der NB: Max! | b |
|---|---|---|---|---|---|
| $a_{1,1}$ | $a_{1,2}$ | … | $a_{1,i}$ | ≤ | $b_1$ |
| $a_{2,1}$ | … | … | … | ≤ | $b_2$ |
| … | … | … | … | ≤ | $b_3$ |
| $a_{n,1}$ | … | … | $a_{n,i}$ | ≤ | $b_n$ |
| $c_1$ | $c_2$ | … | $c_n$ | | |

Das duale Problem lässt sich als Modell und in Tabellenform wie folgt darstellen:

$$Z = \sum_{\substack{n=1 \\ i=1}}^{m} b_n * \gamma_i \rightarrow \text{min!}$$

mit

$$\sum_{\substack{n=1 \\ i=1}}^{m} a_{i,n} * \gamma_i \geq c_n$$

mit $\gamma_i \geq 0$

| $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\gamma_i$ | Art der NB: Min! | c |
|---|---|---|---|---|---|
| $a_{1,1}$ | $a_{1,2}$ | … | $a_{1,i}$ | ≥ | $c_1$ |
| $a_{2,1}$ | … | … | … | ≥ | $c_2$ |
| … | … | … | … | ≥ | $c_3$ |
| $a_{n,1}$ | … | … | $a_{n,i}$ | ≥ | $c_n$ |
| $b_1$ | $b_2$ | … | $b_n$ | | |

Aus dem primalen Maximierungsproblem

G = 7x + 10y und den Nebenbedingungen

$$\begin{aligned} 3x &+ 2y \leq 36 \\ 2x &+ 4y \leq 40 \\ x &+ 0y \leq 10 \\ x, y &\geq 0 \end{aligned}$$

lässt sich das duale Minimierungsproblem

K = $36\gamma_1 + 40\gamma_2 + 10\gamma_3$ mit den Nebenbedingungen

$$\begin{aligned} 3\gamma_1 &+ 2\gamma_2 + \gamma_3 \geq 7 \\ 2\gamma_1 &+ 4\gamma_2 + 0\gamma_3 \geq 10 \\ \gamma_1, \gamma_2, \gamma_3 &\geq 0 \end{aligned}$$

formulieren.

Was sagen nun die $\lambda_i$ aus? Wenn man die „rechte Seite" der Nebenbedingungen im primalen Problem um eine Einheit erhöht, dann erhöht sich der Zielfunktionswert um das $\lambda_i$-fache. Die Schattenpreise geben damit Auskunft darüber, inwiefern Variationen der Nebenbedingungen den Zielfunktionswert beeinflussen.

Es wird nunmehr angenommen, dass sich aus der grafischen Lösung des primalen Problems die Nebenbedingungen 1 und 2 als relevant herausgestellt haben. Dies ist dann auch auf das duale Problem zu übertragen: Wenn die dritte Nebenbedingung irrelevant ist, dann ist $\lambda_3$ = Null. Infolgedessen reduzierten sich die zur Ermittlung der Schattenpreise zu lösenden Gleichungen auf:

$3\gamma_1 + 2\gamma_2 = 7$
$2\gamma_1 + 4\gamma_3 = 10$

Es ergibt sich

$\gamma_1 = 1$
$\gamma_2 = 2$

Bei primalen Minimierungsproblemen handelt es sich bei den Nebenbedingungen regelmäßig um Mindestproduktionsmengen. Bei primalen Maximierungsproblemen handelt es sich grundsätzlich um Kapazitätsrestriktionen.

Wenn sich also im Beispiel die Kapazität in der ersten Nebenbedingung um 6 Einheiten vermindert, so vermindert sich der Zielfunktionswert (= der maximale Gewinn, $b_n$) um $6 * \lambda_1$, also um 6 Geldeinheiten. Dies lässt sich leicht verifizieren, in dem man nunmehr die erste Nebenbedingung entsprechend abändert und den Zielfunktionswert neu berechnet und mit der Ausgangslage vergleicht:

| Ausgangslage | | Variation | |
|---|---|---|---|
| Zielfunktion | max G = 7x + 10y | Zielfunktion | max G = 7x + 10y |
| | 3x + 2y ≤36 | | 3x + 2y ≤30 |
| | 2x + 4y ≤40 | | 2x + 4y ≤40 |
| Ergebnis für x | = 8 | Ergebnis für x | = 5 |
| Ergebnis für y | = 6 | Ergebnis für y | = 7,5 |
| Ergebnis Zielfunktionswert | = 116 | Ergebnis Zielfunktionswert | = 110 |

Analog ergibt sich für Nebenbedingung 2, dass eine Verminderung der Kapazität um eine Einheit den Zielfunktionswert um $1 * \lambda_2 = 2$ vermindert. Vermindert sich die Kapazität in Nebenbedingung 3, so vermindert sich der Zielfunktionswert um $1 * \lambda_3 = 0$, da $\lambda_3 = 0$, also gar nicht. Eine Verminderung der Kapazität würde also ein Verzicht auf Gewinn bedeuten (Opportunitätskosten).

Wenn man die Kapazität um eine Einheit erhöhen würde (könnte), so würde sich der Zielfunktionswert entsprechend erhöhen (Opportunitätsgewinne).

### 3.4.5 Kritische Würdigung der linearen Optimierung

Die Methoden der linearen Optimierung stellen Werkzeuge zur Lösung eines (betriebswirtschaftlichen) Problems für die Entscheidungsvorbereitung bereit. Wie bei allen Modellen besteht auch hier die Schwierigkeit, dieses zu formulieren und dabei einerseits die Komplexität betriebswirtschaftlicher Probleme angemessen zu berücksichtigen und andererseits die Komplexität weitestgehend zu reduzieren ohne die tatsächlichen Gegebenheiten zu verzerren („Optimierung der Modellierung"). Das obige Beispiel zeigt des Weiteren, dass das Verfahren der linearen Optimierung nur bei zwei Variablen anwendbar ist. Die Probleme in der betriebswirtschaftlichen Praxis (Realwelt) stellen letztlich umfassende Optimierungsprobleme dar, die nur schwer in ein Modell gepasst und gesamt optimiert werden können.

## Literatur

Bartels, Hans (1993): Optimierung, lineare. In: Handwörterbuch der Betriebswirtschaft, Teilband 2, 5. Auflage, Spalten 2953-2968. Stuttgart, 1993

*S*chneeweiß, Christoph (1993): Operations Research. In: Handwörterbuch der Betriebswirtschaft, Teilband 2, 5. Auflage, Spalten 2940-2953. Stuttgart, 1993

# Monte Carlo Optimierung

*von Jörg Baumgart und Thomas Holey*

## 1 Einleitung

Eine Kernfrage betriebswirtschaftlicher Entscheidungen ist die Optimierung eines oder mehrerer Ziele. Die Minimierung der anfallenden Kosten oder die Maximierung des zu erzielenden Gewinns sind primäre Ziele. Auf dem Weg hierzu sind in der Regel Teilprobleme in optimaler Weise zu lösen. Nicht erst bei explodierenden Spritpreisen oder einer steigenden LKW-Maut ist die Frage nach dem kürzesten Weg von Interesse. In der schnelllebigen Zeit globaler Märkte mit dem Internet als allgegenwärtiger Plattform gilt es, die Durchlaufzeiten für alle Prozesse zu minimieren.

Unter dem Begriff Operations Research werden die mathematischen Methoden zusammengefasst, die sich mit der Fragestellung befassen, wie die verfügbaren Ressourcen einzusetzen sind, um ein bestimmtes Ziel in optimaler Weise zu erreichen.

Operations Research hat sich seit mehr als einem halben Jahrhundert als eine Disziplin etabliert, in der für bestimmte Problemklassen geeignete Lösungsmethoden entwickelt werden. Generell geht es um die Lösung der folgenden Fragestellung: Welche Werte müssen die (Entscheidungs-)Variablen annehmen, um bei gegebenen Nebenbedingungen eine Zielfunktion zu optimieren, d.h. zu maximieren oder minimieren.

$$Variablen: \vec{x} = (x_1, x_2, ..., x_n)$$
$$Zielfunktion: z = f(\vec{x})$$
$$Nebenbedingungen: g_i(\vec{x}) \leq 0 \quad mit\ i = 1,2,...,m$$

Je nach Gestalt der Zielfunktion oder der Nebenbedingungen ist zwischen linearen und nicht linearen Optimierungsproblemen zu unterscheiden. Die Lineare Optimierung wird mit dem Simplexverfahren sehr gut bewerkstelligt, für Nicht Lineare Optimierung wurde der Lagrangeformalismus entwickelt. Die Tatsache, dass der Definitionsbereich von Variablen auch auf ganze Zahlen beschränkt sein kann, erfordert neue Aspekte bei den Lösungsverfahren, schließlich entfällt die mächtige Methode der Infiniteseminalrechnung für nicht lineare ganzzahlige Optimierungsprobleme. Mit dem Branch & Bound Verfahren, der Dynamischen Optimierung und den Graphenalgorithemen stehen zwar vielfältige Lösungsansätze zur Verfügung, dennoch entziehen sich einige grundlegende Fragestellungen einem analytischen Verfahren oder einem deterministischen Algorithmus zur Bestimmung der Lösung.

Für solche Problemstellungen wurden nicht deterministische Algorithmen entwickelt, die in raffinierter Weise von Zufallsprozessen Gebrauch machen, um näherungsweise eine Lösung zu ermitteln. Der Verwendung von Zufallszahlen, die eben auch beim Glücksspiel von Bedeutung sind, verdanken die im Folgenden betrachteten Lösungsansätze ihren Namen: Monte Carlo Verfahren. Für die Optimierung von besonderem Interesse ist das Verfahren Simulated Annealing. Im zweiten Kapitel wird hierzu eine kurze Einführung gegeben.

Das dritte Kapitel beschäftigt sich mit dem Travelling Salesman Problem (TSP), einer Fragestellung, die eine Komplexität mit sich bringt, die eine exakte Lösung im Allgemeinen unmöglich macht. Exemplarisch für das TSP wird die Vorgehensweise des Simulated Annealing betrachtet und einige Ergebnisse werden vorgestellt.

# 2 Monte Carlo Verfahren

Es gibt zwei prinzipiell verschiedene Arten von Computersimulationen. Bei den so genannten Computerexperimenten wird versucht, die zu untersuchende Problematik möglichst realitätsnah zu modellieren und Lösungsvarianten am Computer zu simulieren. Bei den Monte Carlo Verfahren geht es darum eine mathematische Problemstellung numerisch mit Hilfe von Zufallszahlen zu lösen.

## 2.1 Zufall als Methode

Historisch betrachtet liegen die Anfänge der Monte Carlo Verfahren in der Statistischen Mechanik, jenem Gebiet der Physik, in dem makroskopisch messbare Größen wie die Temperatur oder der Druck eines Gases mit den mikroskopischen Bewegungen der Gasmoleküle in Zusammenhang gebracht werden. Hierbei ist die Berechnung von so genannten Zustandssummen erforderlich, bei der sich die Summation über alle Zustände sämtlicher Moleküle im betrachteten Volumen erstreckt. Vom Physikunterricht erinnert sich mancher vielleicht noch daran, dass die Anzahl der Moleküle nicht gerade gering ist (die Avogadrozahl $6 \cdot 10^{23}$ gibt die Anzahl in ca. 22 Litern an bei Normalbedingungen).
Die Idee bestand nun darin, nicht über alle möglichen Zustände zu summieren, sondern nur über einige, zufällig ausgewählte. Bei einer reinen Zufallswahl ist dieses Verfahren nicht erfolgreich, da die Zahl der Möglichkeiten viel zu groß ist. Berücksichtigt man bei der Auswahl aber, dass den Zuständen eine Wahrscheinlichkeitsverteilung (die Boltzmann-Verteilung) zu Grunde liegt und gewichtet die Zustände gemäß dieser Wahrscheinlichkeitsverteilung, ergeben sich schon mit relativ wenigen Zuständen erstaunlich gute Ergebnisse. Dieses Vorgehen wird als ‚importance sampling' bezeichnet. Die Methode der Monte Carlo Simulation war hiermit geboren [Met1953].

Monte Carlo Verfahren sind inzwischen zu einer wichtigen Methodik innerhalb der Numerischen Mathematik geworden. So werden entsprechende Verfahren auch zur näherungsweisen Berechnung von mehrdimensionalen bestimmten Integralen eingesetzt.
Bevor die Anwendung auf Optimierungsprobleme genauer betrachtet wird, sei noch kurz auf die Zufallszahlen eingegangen, die den Monte Carlo Verfahren zu Grunde liegen.

Eine Zufallszahlenfolge liegt vor, wenn die Folge von Zahlen keinerlei Bildungsgesetzmäßigkeit aufweist. Zwischen unterschiedlichen Zahlen der Folge darf keinerlei Korrelation nachweisbar

sein. Solche zufälligen Zahlenfolgen lassen sich im Computer prinzipiell über Hardwareparameter generieren. Die n-te Nachkommastelle der an einem integrierten Schaltkreis anliegenden Spannung unterliegt mit Sicherheit zufälligen Schwankungen. Solche hardwaregenerierten Zufallszahlen können allerdings den Monte Carlo Algorithmen nicht schnell genug zur Verfügung gestellt werden. Daher werden Algorithmen eingesetzt, die Zufallszahlenfolgen generieren. Dieser Widerspruch in sich löst sich folgendermaßen auf: Algorithmen generieren Zahlenfolgen, die sich erst nach einer sehr großen Periodenlänge wiederholen. Innerhalb einer Periode liegen tatsächlich keine Korrelationen zwischen den generierten Zahlen vor. Solche Zahlenfolgen werden als Pseudozufallszahlen bezeichnet. Solange in der Simulation weniger Zufallszahlen benötigt werden als einer Periode des Pseudozufallsgenerators zu Grunde liegen, treten keine Probleme auf.

Mit der multiplikativen Kongruenz [Leh1949] und der Shiftregister Methode stehen sehr leistungsstarke Algorithmen zur Verfügung, die Pseudozufallszahlen mit genügend hoher Periodenlänge generieren. Die überaus hohen Periodenlängen beruhen auf zahlentheoretischen Eigenschaften der Mersenne Exponenten [Kir1981]. Die gängigen Programmiersprachen bieten in der Regel intrinsische Pseudozufallszahlen an, die meist auf der multiplikativen Kongruenz beruhen. Eine Überprüfung der Periodenlänge im Hinblick auf die im Rahmen der Simulation anfallenden Zufallszahlen ist grundsätzlich zu empfehlen. So gut die Ergebnisse von Monte Carlo Verfahren auch sein können, solange die Zufälligkeit der Zufallszahlen sichergestellt ist, wenn die Periodenlänge des Pseudozufallsgenerators überschritten wird oder andere Formen der Korrelation vorliegen, sind die Ergebnisse wertlos.

Andersherum werden Pseudozufallszahlgeneratoren getestet, indem geprüft wird, wie gut Monte Carlo Simulationen die Ergebnisse exakt berechenbarer Modelle reproduzieren (ein Beispiel hierfür ist das zweidimensionale Isingmodell).

## 2.2 Optimierung mit Simulated Annealing

Ein heute noch wichtiges und historisch gesehen das erste Monte Carlo Verfahren zur Lösung von Optimierungsproblemen heißt Simulated Annealing [Kir1983]. Der Begriff ist wie Vieles in diesem Umfeld der Physik entnommen, da die Verfahren zunächst für physikalische Fragestellungen entwickelt wurden, ehe der Nutzen für die Lösung betriebswirtschaftlicher Problemstellungen erkennbar war.
Um besonders reine Kristalle (Metalle oder künstliche Diamanten) herstellen zu können wird eine Schmelze (ein weitgehend ungeordneter Zustand weit weg vom Energieminimum) sehr langsam abgekühlt. Auf diese Weise wird die Bildung von Fehlstellen weitgehend vermieden und ein reiner Kristallzustand, der energetisch betrachtet minimal ist, kann sich ausbilden.

Dieser Vorgang wird nun zur Lösung eines Optimierungsproblems simuliert. Auf der Suche nach dem Minimum (Maximierungsprobleme lassen sich immer analog betrachten) beginnt man bei einer beliebig schlechten Lösung und führt schrittweise kleine Veränderungen durch. In Anlehnung an die Terminologie der Genetik werden die Veränderungen auch als Mutationen bezeichnet.

Wird durch eine Mutation eine bessere Lösung generiert, so wird diese akzeptiert. Würde eine schlechtere Lösung immer verworfen, so fände das Verfahren relativ schnell ein lokales Extremum, aus dem es dann allerdings kein Entrinnen mehr gäbe.

Erst wenn auch hin und wieder Verschlechterungen in einer geeigneten Größenordnung akzeptiert werden, besteht die Chance, lokale Extrema zu überwinden. Eine generelle Akzeptanz von Verschlechterungen ist dagegen ebenso wenig hilfreich, da dieses Vorgehen einem reinen Ausprobieren gleichkommt, was bei der großen Zahl von möglichen Lösungen in aller Regel nicht Ziel führend ist. Die fol-

gende Abbildung veranschaulicht diese Suche nach einem globalen Minimum für eine Funktion mit nur einer Variablen.

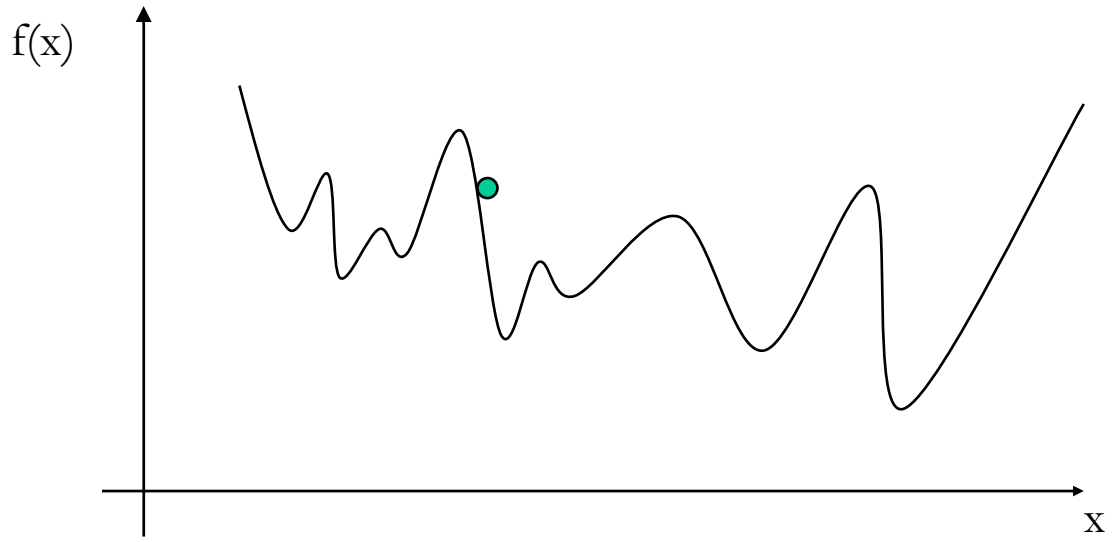



Abb.1: Funktion mit vielen lokalen Extrema

Beim Simulated Annealing Verfahren gilt es, die Akzeptanz von Verschlechterungen mit einer geeigneten Wahrscheinlichkeit abzubilden. Diese Wahrscheinlichkeit soll einerseits mit dem Ausmaß der Verschlechterung eines Zielfunktionswertes skalieren, andererseits soll das Akzeptanzniveau insgesamt im Laufe der Simulationszeit abnehmen. Eine Wahrscheinlichkeitsfunktion, die dieses Verhalten für ein Minimierungsproblem beschreibt, lässt sich folgendermaßen modellieren:

$\vec{x}_a$ : *aktuelle Lösung*

$\vec{x}_n$ : *neue Lösung*

$z = f(\vec{x})$ : *Zielfunktion*

$\Delta z = f(\vec{x}_n) - f(\vec{x}_a)$

$\beta$ : *Parameter* $> 0$

$$p(\Delta z, \beta) = \begin{cases} 1 & \text{für } \Delta z \le 0 \\ \exp(-\beta \cdot \Delta z) & \text{für } \Delta z > 0 \end{cases}$$

Verbesserungen werden immer (mit Wahrscheinlichkeit 1) akzeptiert, Verschlechterungen dagegen mit einer abnehmenden Wahrscheinlichkeit. Der Parameter β wird im Laufe der Simulation allmählich erhöht, er charakterisiert hier die Simulationszeit und steht im oben beschriebenen physikalischen Abkühlprozess für die abnehmende Temperatur. In der folgenden Abbildung ist der Verlauf der Akzeptanzwahrscheinlichkeit für verschiedene Werte von β beschrieben.

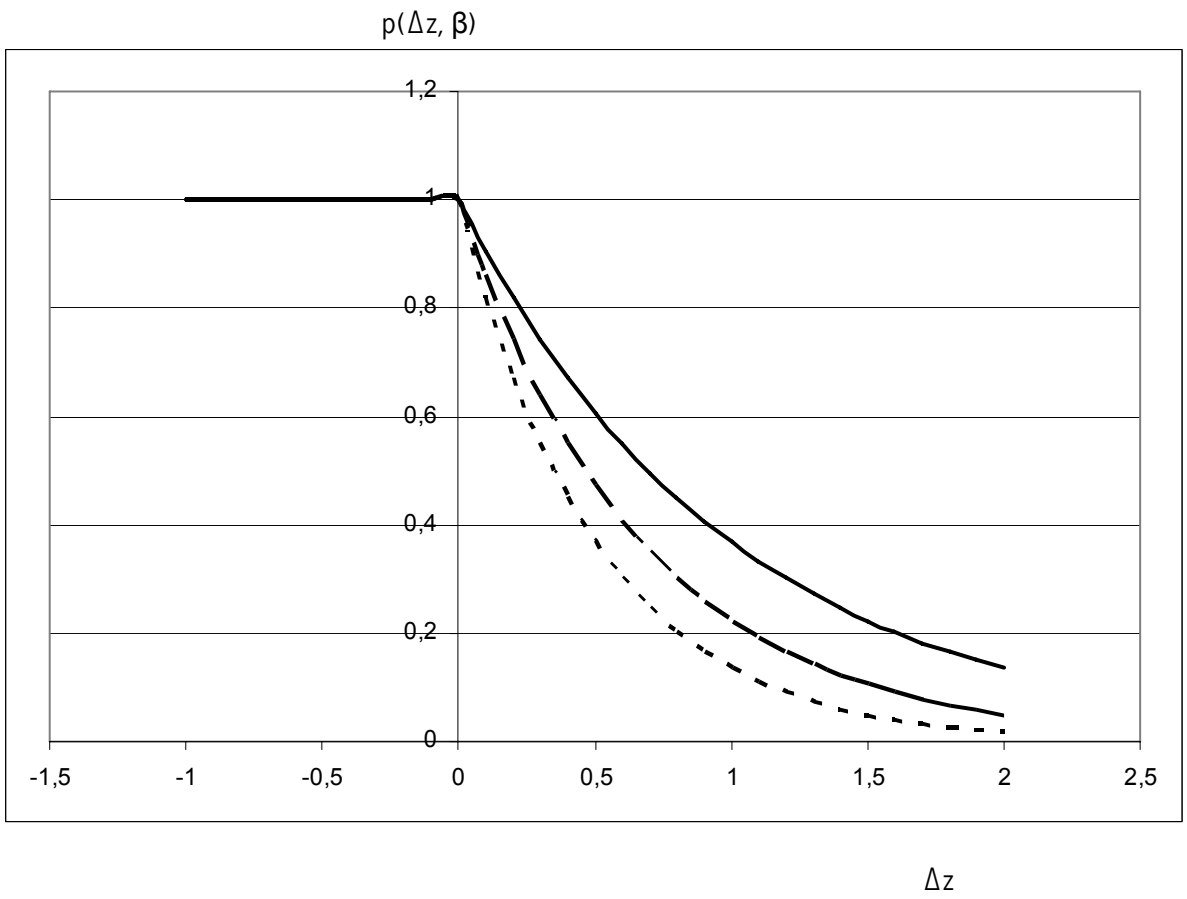



Abb.2: Akzeptanzwahrscheinlichkeit für verschieden Werte von β

## 2.3 Der Simulated Annealing Algorithmus

In diesem Abschnitt wird beschrieben, wie das in 2.2 geschilderte Verfahren in einem Algorithmus umgesetzt werden kann.

Der Parameter β wird auf einen hinreichend kleinen (siehe unten) Anfangswert gesetzt. Ausgehend von einer beliebigen Startlösung wird eine zufällige Veränderung vorgenommen, dann wird die Differenz der Zielfunktionswerte beider Lösungen verglichen und die entsprechende Akzeptanzwahrscheinlichkeit berechnet. Eine Lösung mit schlechterem Zielfunktionswert wird mit eben dieser Wahrscheinlichkeit akzeptiert, dies ist das nicht deterministische Element des Algorithmus.

Erreichen lässt sich diese bedingte Akzeptanz, indem der Wert $p(\Delta z, \beta)$ mit einer Zufallszahl RN aus dem Intervall [0, 1] verglichen wird. Ist $RN < p(\Delta z, \beta)$, so wird die Veränderung akzeptiert, ansonsten wird sie verworfen. Die vor der Veränderung liegende Lösung ist dann Ausgangspunkt für weitere Simulationsschritte.

In dieser Weise wird eine gewisse Anzahl von Schritten ausgeführt, bevor man eine geringfügige Erhöhung des Parameters β vornimmt und dann wieder einige Schritte durchführt. Kompakt geschrieben in einer Pseudocode-Schreibweise sieht der Simulated Annealing Algorithmus folgendermaßen aus:

1. *Wähle beliebige Startlösung* $\vec{x}_a$
2. *Wähle für* $\beta$ *eine kleine positive Zahl*
3. *Verändere die aktuelle Lösung* $\vec{x}_a$ *geringfügig* $\rightarrow \vec{x}_n$
4. *Berechne* $\Delta z$ *und* $p(\Delta z, \beta)$
5. *Wähle Zufallszahl* $RN \in [0,1]$
6. *Für* $RN \leq p(\Delta z, \beta)$ $\vec{x}_a = \vec{x}_n$
   *Sonst Lösche* $\vec{x}_n$
7. *Gehe* $(n)$ *mal zu* 3.
8. *Erhöhe* $\beta$
9. *Gehe* $(m)$ *mal zu* 3.

Die Steuerung des Parameters β kann in einfacher Weise so erfolgen, dass β bei jedem Durchlauf um einen geringen Betrag oder Prozentsatz erhöht wird. Ein geeignet kleiner Anfangswert für β ergibt sich, wenn der überwiegende Teil von Verschlechterungen zunächst akzeptiert wird.

Die Qualität des Ergebnisses, das für die Optimierung erzielt werden kann, hängt zum einen davon ab wie viele Schritte überhaupt ausgeführt werden, zum anderen aber auch davon, bei dem jeweiligen Wert von β eine ‚genügend hohe' Anzahl von Schritten durchzuführen. Wie viele das sind, kann in einem ersten Ansatz an Hand der Verbesserung des erzielten Ergebnisses ermittelt werden.

Eine wissenschaftlich fundierte Vorgehensweise besteht darin, den Verlauf des Parameters β und die Anzahl der jeweils durchgeführten Simulationsschritte über die Analyse der Fluktuationen, die sich für die Zielfunktionswerte ergeben, zu steuern [Mit1985]. Die hier angewandten Mechanismen beruhen auf Erkenntnissen aus der Statistischen Mechanik bei der Beschreibung von Phasenübergängen.

Bevor der Algorithmus auf die Lösung des Travelling Salesman Problems im dritten Kapitel angewandt wird, seien zunächst noch die Möglichkeiten erörtert, wie Restriktionen berücksichtigt werden können.

## 2.4 Berücksichtigung von Restriktionen

Die betriebswirtschaftlichen Rahmenbedingungen stellen bei einem Optimierungsproblem die Restriktionen bzw. Nebenbedingungen dar. Werden wie in Kapitel 3 dargestellt optimale Touren gesucht, so sind denkbare Restriktionen Zeitfenster, in denen die Orte einer Tour aufzusuchen sind, oder Einschränkungen bei der Verwendung von Fahrzeugen bis hin zu der Einhaltung gesetzlicher Rahmenbedingungen für Fahrt- und Pausenzeiten.

Für die Berücksichtigung von Restriktionen im Rahmen des Simulated Annealing Verfahrens gibt es zwei prinzipielle Ansätze. Zum einen besteht die Möglichkeit, zufällige Veränderungen an den Lösungen nur in einer Art und Weise vorzunehmen, dass der Raum der zulässigen Lösungen nicht verlassen wird. Damit ist sichergestellt, dass auch das ermittelte Optimum auf einer zulässigen Lösung beruht. Ein Nachteil dieser Vorgehensweise kann sich ergeben, wenn

durch die Einhaltung der Restriktionen die Änderungen der Zielfunktion, die sich auf Grund der Mutationen ergeben, immer sehr groß sind. Dann liegt bildlich gesprochen ein im Vergleich zum restriktionsfreien Optimierungsproblem ein wesentlich stärker zerklüfteter Lösungsraum vor, und das oben beschriebene Überspringen von kleineren Barrieren im Laufe zunehmender Simulationszeit verliert an Bedeutung.

Ein anderer Ansatz, die Restriktionen zu berücksichtigen, lehnt sich an den Lagrangeformalismus an [Neu1993]. Die Zielfunktion wird durch ‚Strafterme' ergänzt, die bei Verletzung einer Restriktion einen ungewünschten Beitrag liefern. Die Gewichtsfaktoren der ‚Strafterme' sind die Lagrange-Multiplikatoren. Für ein Minimierungsproblem sieht die Betrachtung folgendermaßen aus.

$$Zielfunktion: z = f(\vec{x})$$
$$Nebenbedingungen: g_i(\vec{x}) \le 0 \quad mit\ i = 1,2,...,m$$
$$L: \Re^{n+m} \to \Re$$
$$L(\vec{x},\vec{u}) = f(\vec{x}) - \sum_{i=1}^{m} u_i g_i \ mit$$
$$u_i \ge 0: Lagrange - Multiplikatoren$$

Jede Verletzung einer Nebenbedingung ( $g_i(\vec{x}) > 0$ ) bedeutet dann im Sinne der Minimierung eine Verschlechterung. Über die Steuerung der Lagrangeparameter muss jetzt dafür gesorgt werden, dass die Einhaltung einer Restriktion nicht belohnt wird und dass im Laufe der Simulationen die Verletzung von Restriktionen immer stärker bestraft wird.

Insgesamt gesehen ist mit dieser Herangehensweise ein eher heuristischen Ansatz verbunden, die neu hinzukommenden Parameter zu steuern, dafür lassen sich aber Restriktionen in sehr vielfältiger Weise abbilden.

# 3 Das Travelling Salesman Problem

Das Travelling Salesman Problem (TSP) stellt ein grundlegendes Optimierungsproblem dar. Im Rahmen der Klassifikation der Komplexitätstheorie gilt es als ein NP-Problem (non-polynominal), da die Zahl der Möglichkeiten nicht polynomial von der Systemgröße abhängt. Genauer gesagt gehört das TSP zu den NP-vollständigen Probleme [Gar1978].

## 3.1 Problemstellung

Wie der Name des Travelling Salesman Problems schon ausdrückt geht es um ein Optimierungsproblem eines Handlungsreisenden. Ausgehend von seinem Startort hat er alle Orte seiner Kunden aufzusuchen und kehrt schließlich zum Startort zurück. Allein die Reihenfolge, in

der die Kunden aufgesucht werden, bleibt ihm überlassen. Die Reihenfolge entscheidet über die zurückgelegte Wegstrecke, die es zu minimieren gilt.

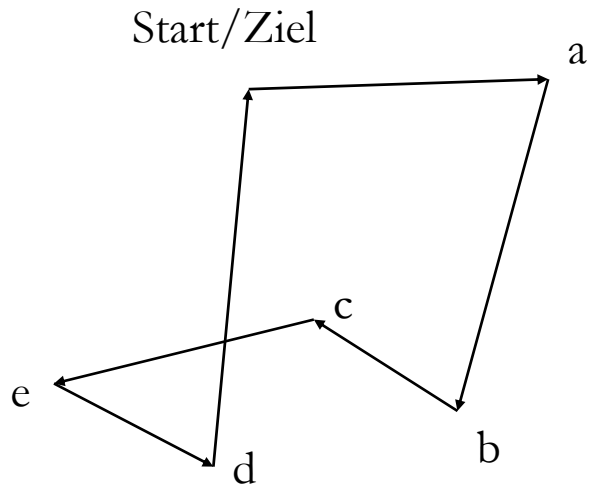



Abb.3: Das Travelling Salesman Problem

Bei einigen wenigen Orten lässt sich das Optimum durch Ausprobieren rasch finden.
Da die Zahl der Möglichkeiten, die Reihenfolge der n aufgesuchten Kundenorte zu variieren, aber n! (n Fakultät) ist und damit nicht polynomial, versagt das Ausprobieren schon für n > 10.

Ein stupider Algorithmus, der alle Möglichkeiten durchprobieren würde (brute force), scheitert selbst bei der heute erstaunlichen Rechnerleistung (mit Taktraten unter einer Nanosekunde). Schon ab ca. dreißig Orten übersteigt die Rechenzeit das geschätzte Alter des Universums von 14 Milliarden Jahren.

## 3.2 Lösung durch Simulated Annealing

In diesem Abschnitt wird betrachtet, wie für das TSP mit dem Verfahren Simulated Annealing eine gute Näherungslösung erzielt werden kann. Der Kern des in 2.3 beschriebenen Algorithmus ist generischer Natur, d.h. er ist unabhängig vom eigentlichen Optimierungsproblem. Auf die eigentliche Problemstellung bezogen sind die Fragen, was eine Lösung ist, wie man diese verändert und welche Auswirkung dies auf die Zielfunktion hat.

Eine Lösung für das TSP wird im Folgenden als Tour bezeichnet, sie muss im Startpunkt beginnen, dort auch wieder enden und ansonsten muss jeder Ort in einer Tour genau einmal aufgesucht werden.

Die Darstellung einer Tour ist besonders einfach, sie lässt sich als Folge der aufgesuchten Orte schreiben. Für die Tour in Abbildung 3 ergibt sich also: T1 = {Start, a, b, c, e, d, Start}, wobei der Start auch das Ziel darstellt und in der Liste unveränderbar ist. Die kleinste mögliche Veränderung einer Tour ergibt sich durch den so genannten Lin-2-Opt Schritt. Hierbei werden (zufällig) zwei Verbindungen der Tour herausgegriffen und die dabei entstehenden offenen Enden

so geschlossen, dass sich die Reihenfolge der neuen Tour von der alten unterscheidet, wie dies in Abbildung 4 dargestellt ist. Zunächst werden die Verbindungen c-->e und d-->Start entfernt, dann wird die Tour die Einfügen von c-->d und e-->Start wieder geschlossen. Die neue Tour T2 lautet also T2 = {Start, a, b, c, d, e, Start}.

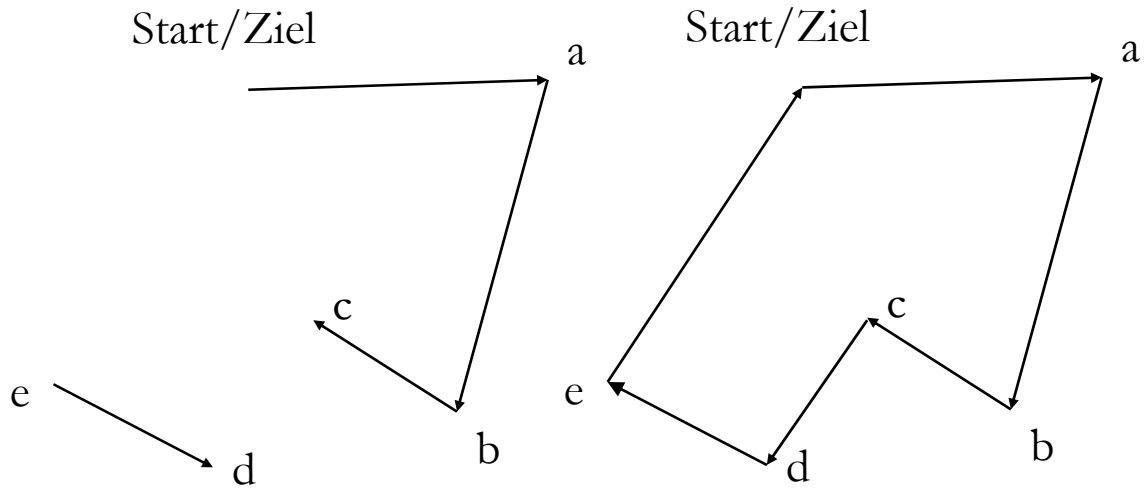



Abb.4: Lin – 2- Opt Schritt

Die Zielfunktionsveränderung ist leicht zu bestimmen, hier handelt es sich ja um die Längendifferenz der beiden zu vergleichenden Touren.

Dieser elementare Schritt ist nun gemäß des Algorithmus 2.3 wiederholt auszuführen, wobei die zu tauschenden Kanten zufällig gewählt werden und die Entscheidung der Akzeptanz gemäß des Simulated Annealing Verfahrens durchzuführen ist.

Im folgenden Kapitel werden einige Ergebnisse für ein sehr bekanntes Benchmark-TSP mit 442 Orten vorgestellt [Grö1991].

## 3.3. Simulationsergebnisse

Für den Test der Qualität von Näherungsverfahren zur Lösung von TSPs sind exemplarische Problemstellungen, für die die Lösung z.B. auf Grund spezieller Symmetrien bekannt ist, von großer Bedeutung. Ein solches Problem ist das oben zitierte Problem mit 442 Orten. Die optimale Lösung hat eine Länge von 50 680.

Im Folgenden wird betrachtet, welche Lösungen das Verfahren Simulated Annealing liefert.

Als Parameter der Simulation sind zu betrachten:

- Anfangswert für β
- Der Parameter α, der die Erhöhung von β steuert
- Die Anzahl der inneren Schleifendurchläufe, das sind die Simulationsschritte bei jeweils einem Wert von β.
- Die Anzahl der äußeren Schleifendurchläufe, das ist die Anzahl der Schritte, in denen β erhöht wird.

Die Erhöhung von β erfolgt hier gemäß:

$$\beta \to \beta \cdot \frac{1}{1-\alpha}$$

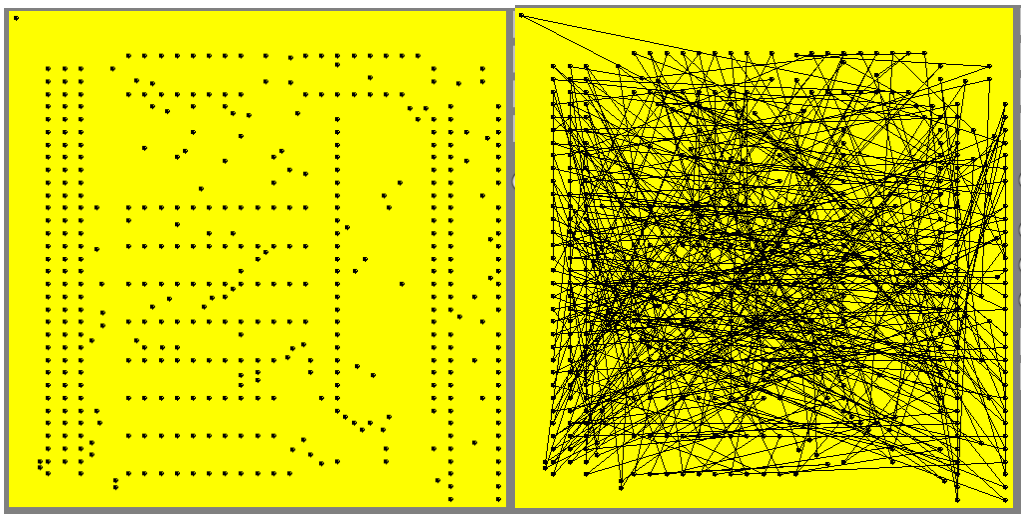


Abb.5: Das 442-Orte TSP

Abb.6: Eine beliebige Ausgangslösung

Die Abbildung 5 zeigt das Ausgangsproblem also die Lage der 442 Orte durch die eine optimale Tour zu finden ist.

Die Anfangstour für das Näherungsverfahren kann beliebig sein, muss lediglich die Bedingung erfüllen, dass es sich um eine geschlossene Rundtour handelt, in der jeder Ort einmal aufgesucht wird. Die in Abb. 6 dargestellte Tour erfüllt diese Bedingung, wenn dies auch optisch nicht nachvollziehbar ist.

Für α = 0,001 und β = 0,001 ergibt sich mit 500 000 Durchläufen der inneren Schleife und ca. 8 000 Durchläufen der äußeren Schleife eine Tourlänge von 51150 (vgl. Abb. 7). Dies stellt eine Abweichung von ca. 1% von der für dieses Problem bekannten, exakten Lösung dar.

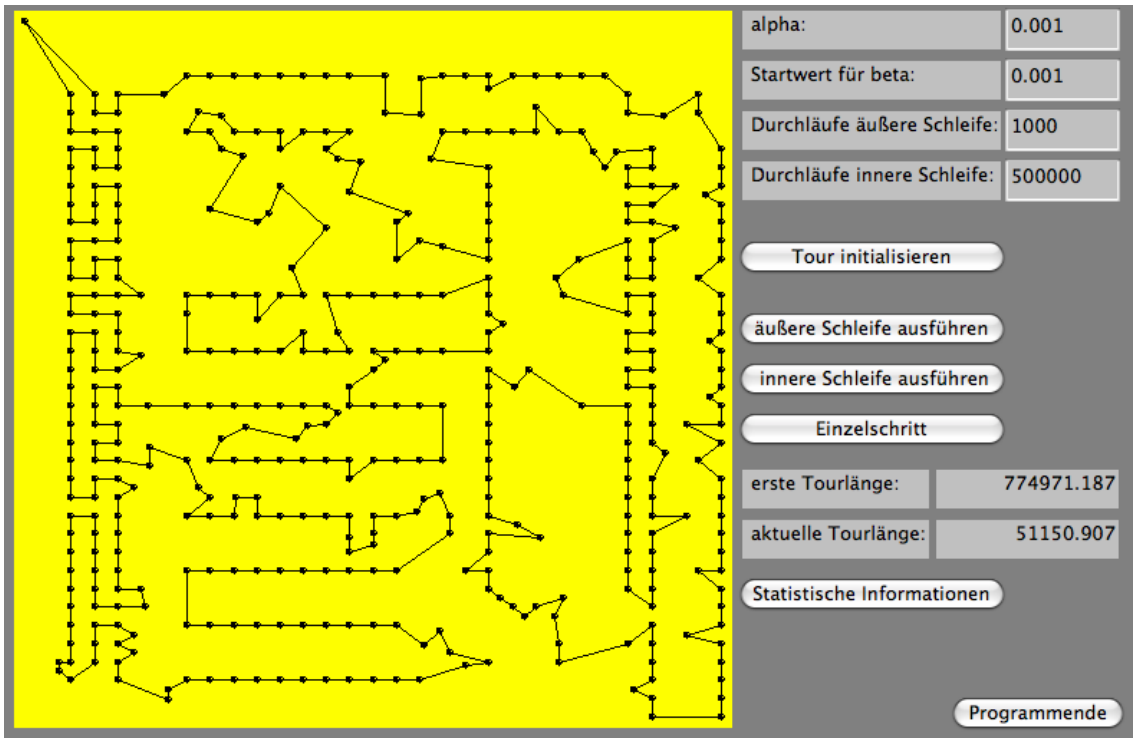



Abb.7: Näherungslösung für das 442-Orte TSP

Waren solche Simulationen vor wenigen Jahren noch auf Großrechner beschränkt, so lassen sich diese Ergebnisse heute bereits mit einigen Minuten Rechenzeit auf leistungsstarken PCs erzielen.

## 4 Zusammenfassung

Simulated Annealing ist ein Monte Carlo Verfahren zur Lösung von Optimierungsproblemen, wie sie sich unter anderem im betriebswirtschaftlichen Umfeld stellen.

Der generische Algorithmus lässt sich mit vertretbarem Aufwand mit einer problemspezifischen Modifikation der Lösungen im Lösungsraum kombinieren.

Am Beispiel des Travelling Salesman Problems wird die Funktionsweise des Algorithmus erläutert und dessen Leistungsfähigkeit an Hand eines 442-Ort TSP demonstriert.

## Literatur

[Gar1978]: Garey, Michael R.; Johnson, David S. (1978): Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, San Francisco, 1978.

[Grö1991]: Grötschel, M., Holland, O. (1991): Solution of Large Symmetric TSP, Math. Programming 51, pp 141 -202, 1991

[Leh1949]: Lehmer, D. H. (1949): Mathematical methods in large-scale computing units Proceedings 2nd Symposium of Large-Scale Digital Calculating Machinery, 1949 .

[Kir1981]: Kirkpatrick, S.; Stoll, E.; Comp., J. (1981): Phys. 40, pp 517 – 531, 1981

[Kir1983]: Kirkpatrick, S. et al. (1983): Optimization by Simulated Annealing, Science Vol.220, pp 671-680, 1983

[Mer1998]: Matsumoto, M.; T. Nishimura, T. (1998): Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans. on Modeling and Computer Simulations, 1998

[Met1953]: Metropolis, N. et al., J. Chem. Phys. Vol.21 pp 1087 – 1092, 1953

[Mit1985]: Mitra, D. et al. (1985): Convergence and Finite Time Behaviour of SA in Proc. 24th Conf. on Decision and Control pp 761 – 767, 1985

[Neu1993]: Neuer, K.; Marlock, M. (1993): Operations Reserach, Hanser 1993