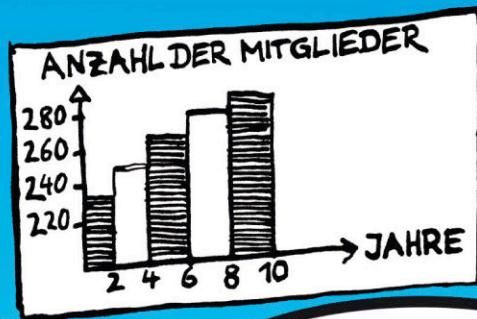


ANDREAS LINDENBERG, IRMGARD WAGNER
ILLUSTRIERT VON PETER FEJES

STATISTIK

macchiato

CARTOONKURS FÜR SCHÜLER UND STUDENTEN



TRAU KEINER
STATISTIK...

... DIE DU NICHT
DURCHSCHAUST...



2. Auflage

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Die Informationen in diesem Buch werden ohne Rücksicht auf einen eventuellen Patentschutz veröffentlicht. Warennamen werden ohne Gewährleistung der freien Verwendbarkeit benutzt. Bei der Zusammenstellung von Texten und Abbildungen wurde mit größter Sorgfalt vorgegangen. Trotzdem können Fehler nicht ausgeschlossen werden. Verlag, Herausgeber und Autoren können für fehlerhafte Angaben und deren Folgen weder eine juristische Verantwortung noch irgendeine Haftung übernehmen. Für Verbesserungsvorschläge und Hinweise auf Fehler sind Verlag und Herausgeber dankbar.

Alle Rechte vorbehalten, auch die der fotomechanischen Wiedergabe und der Speicherung in elektronischen Medien. Die gewerbliche Nutzung der in diesem Produkt gezeigten Modelle und Arbeiten ist nicht zulässig.

Fast alle Produktbezeichnungen und weitere Stichworte und sonstige Angaben, die in diesem Buch verwendet werden, sind als eingetragene Marken geschützt. Da es nicht möglich ist, in allen Fällen zeitnah zu ermitteln, ob ein Markenschutz besteht, wird das ® Symbol in diesem Buch nicht verwendet.

10 9 8 7 6 5 4 3 2 1

13 12 11

ISBN 978-3-86894-078-7

© 2011 Pearson Studium

ein Imprint der Pearson Education Deutschland GmbH

Martin-Kollar-Str. 10-12, D-81829 München

Alle Rechte vorbehalten

www.pearson-studium.de

Programmleitung: Birger Peil, bpeil@pearson.de

Fachlektorat: Dr. Ludwig Feichtinger, Universität Salzburg

Korrektorat: Petra Kienle, Fürstenfeldbruck

Herstellung: Martha Kürzl-Harrison, mkuerzl@pearson.de

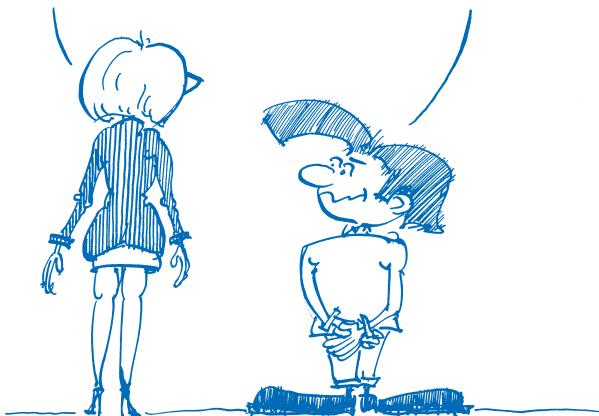
Satz: m2 design, Sterzing, www.m2-design.org

Druck und Verarbeitung: Bercker, Kevelaer

Printed in Germany

WAS GLAUBEN SIE? BEI WEM VON UNS BEIDEN IST DIE WAHRSCHEINLICHKEIT GRÖSSER, DAS RENTENALTER ZU ERREICHEN?

WIE KOMMEN SIE JETZT AUF DIESE FRAGE? ABER OHNE IHMEN ZU NAHE TREten ZU WOLLEN, SIE HABEN DIE BESSERE BEDINGTE WAHRSCHEINLICHKEIT DAFÜR.

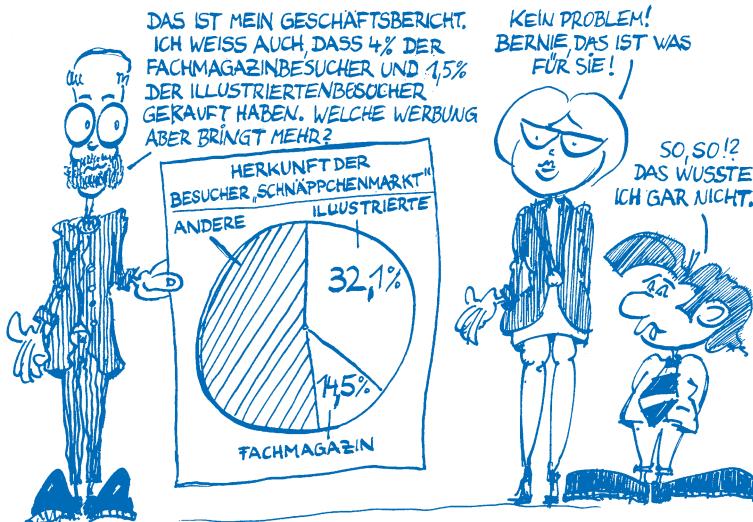


Beispiel mit bedingter Wahrscheinlichkeit

Statistica legt Bernie den Geschäftsbericht der Firma „Schnäppchenmarkt“ auf den Tisch. „Schnäppchenmarkt“ bietet seine Produkte im Internet über ein Shoppingsystem an und lobt in dem Geschäftsbericht seine verkaufsfördernden Maßnahmen. Anhand eines Kreisdiagramms wird der Erfolg einer Werbeaktion des letzten Jahres veranschaulicht.

Auf zwei gut frequentierten Webseiten eines Fachmagazins und einer Illustrierten wurde jeweils ein Werbebanner platziert, der dem „Schnäppchenmarkt“ Kunden zuführen sollte. Aus dem Kreisdiagramm lässt sich ablesen, dass 14,5% aller Online-Besucher des „Schnäppchenmarkts“ über das Fachmagazin kamen und 32,1% über die Illustrierte. Außerdem verrät der Geschäftsbericht, dass 4% der Fachmagazinbesucher und 1,5% der Besucher, die über die Illustrierte kamen, tatsächlich gekauft haben.

Der Geschäftsbericht sagt aber nichts darüber aus, welcher der beiden Werbeplätze der erfolgreichere war. Genau dies soll Bernie für Statistica ermitteln.



Ganz offensichtlich wird die Wahrscheinlichkeit, mit der ein Besucher der Internetplattform auch tatsächlich kauft, von einem anderen Ereignis beeinflusst, nämlich dadurch, über welchen Werbeplatz er zum „Schäppchenmarkt“ geleitet wurde. Diese Ereignisse sind dann voneinander abhängig.

Wir legen drei Ereignisse F , I und U fest:

F : Der Besucher kommt über die Werbung im Fachmagazin.

I : Der Besucher kommt über die Werbung in der Illustrierten.

U : Übrige Besucher

Außerdem unterscheiden wir noch die Käufer und die Nicht-Käufer:

K : Besucher **kauf**t ein Produkt.

\bar{K} : Besucher **kauf**t **nicht**.

Mit welcher Wahrscheinlichkeit kommt nun ein Besucher über das Fachmagazin zum „Schnäppchenmarkt“ **und** kauft ein Produkt?

Wir fragen also nach $P(F \cap K)$. Laut Auswertung von Bernie ist $P(F) = 14,5\%$ die Wahrscheinlichkeit dafür, dass ein Besucher über das **Fachmagazin** auf die Seiten des „Schäppchenmarkts“ kommt.

Von allen Besuchern des Fachmagazins werden 4 % ein Produkt kaufen, also gilt:

$$P(K|F) = 4\%.$$

$P(K|F)$ ist die abkürzende Schreibweise für $P(K$ unter der Bedingung, dass F eingetreten ist).

Gesucht ist die Wahrscheinlichkeit, dass jemand über das **Fachmagazin** gekommen ist **und** gleichzeitig **Käufer** ist, also $P(F \cap K)$. Das können wir mit dem Multiplikationsgesetz ausrechnen. Nun ist

$$P(F \cap K) = P(K|F) \cdot P(F) = 0,04 \cdot 0,145 = 0,0058$$

die gesuchte Antwort, in Prozent 0,58%, was plausibel erscheint, denn schließlich kaufen 4% der Fachmagazinbesucher, welche wiederum 14,5% von allen Besuchern ausmachen.

Von allen Besuchern des „Schnäppchenmarkts“ kommen

$$P(I) = 32,1\%$$

über die **Illustrierte**. Von diesen wiederum kaufen

$$P(K|I) = 1,5\%$$

tatsächlich ein Produkt. Damit ist also

$$P(I \cap K) = P(K|I) \cdot P(I) = 0,48\%$$

die Wahrscheinlichkeit dafür, dass ein zufällig ausgesuchter Besucher über die **Illustrierte** kam **und** ein Produkt **gekauft** hat.

Falls also die Werbeplätze bei der Illustrierten und dem Fachmagazin gleich viel kosten, muss die Anzeige beim Fachmagazin eindeutig als erfolgreicher eingestuft werden, obwohl in absoluten Zahlen gemessen weniger Besucher über das Fachmagazin kommen.

Zur Behandlung abhängiger und unabhängiger Ereignisse fassen wir zusammen:

Sind zwei Ereignisse A und B voneinander **unabhängig**, so gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

Ist dagegen B von A **abhängig**, so gilt:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

allgemeines Multiplikationsgesetz

Das zweite Multiplikationsgesetz ist das allgemeine. Sind nämlich B und A unabhängig, dann gilt

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

da die Wahrscheinlichkeit für das Eintreten von B nicht durch das Eintreten von A beeinflusst wird.



Mit diesen Überlegungen sind wir nun gerüstet für unseren Kriminalfall auf hoher See. Dabei hilft uns der Nachlass eines englischen Mathematikers aus dem Jahre 1763.

Satz von Bayes

Herr Bernie und Statistica sind auf einem Kreuzfahrtschiff unterwegs und gönnen sich ein paar Tage Urlaub. An einem Nachmittag wird in der Kabine von Diva Dada Damur eingebrochen und eine wertvolle Uhr entwendet. Der Schiffsdetektiv ermittelt. Nach einigen Befragungen bleiben zwölf Verdächtige übrig, unter denen sich auch der Dieb befindet. Wer aber war es?

Der Detektiv möchte zur endgültigen Aufklärung einen Lügendetektor einsetzen. Den Schuldigen erkennt das Gerät mit einer Verlässlichkeit von 92%, einen Unschuldigen sogar mit einer Verlässlichkeit von 98%. – Statistica rät dem Detektiv dennoch von dem Einsatz des Lügendetektors ab. Der Detektiv ist überrascht. Sprechen diese Werte etwa nicht für sich?



Bernie soll berechnen, mit welcher Wahrscheinlichkeit einer der zwölf Verdächtigen schuldig getestet wird, obwohl er unschuldig ist. Dazu definiert er zunächst einmal folgende Ereignisse:

S: Der Verdächtige ist **schuldig**.

U: Der Verdächtige ist **unschuldig**.

D: Das Ergebnis des Tests am Lügendetektor ist positiv.

U ist offensichtlich das Gegenereignis zu *S*. Wenn ein Verdächtiger ausgewählt wird, weiß man nicht, ob man den Schuldigen erwischt hat, ob also *S* oder *U* eingetreten ist. Dies lässt sich nur mutmaßen, d.h. mit einer bestimmten Wahrscheinlichkeit aussagen. *S* und *U* sind daher zunächst **hypothetische Ereignisse** oder einfach **Hypothesen**.

Wenn der Detektor ausschlägt, ist das Ereignis *D* dagegen eingetreten und wir suchen die Wahrscheinlichkeit für *S*, dass der Verdächtige nun wirklich schuldig ist.

Wir fragen also nach der **bedingten Wahrscheinlichkeit**

$$P(S|D),$$

der Wahrscheinlichkeit für *Schuldig*, unter der Bedingung, dass der *Detektor* ausschlägt.

Zunächst einmal sind uns folgende Wahrscheinlichkeiten bekannt:

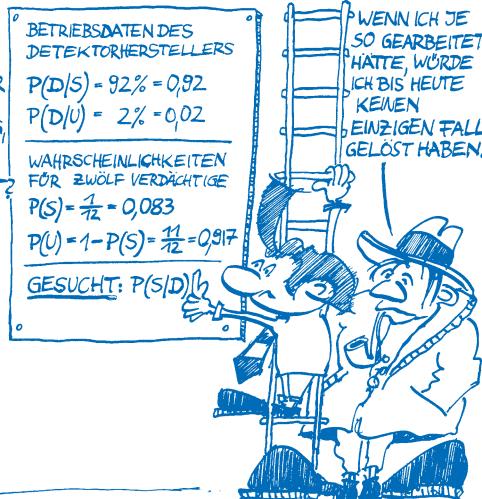
Unter den zwölf Verdächtigen gerade den Schuldigen auszuwählen, geschieht mit der Wahrscheinlichkeit

$$P(S) = \frac{1}{12} = 0,083.$$

Entsprechend ist die Wahrscheinlichkeit, einen Unschuldigen auszuwählen

$$P(U) = 1 - P(S) = \frac{11}{12} = 0,917.$$

DIE GEFÄHR IST IMMER,
DASS EIN UNSCHULDIGER
BÜSST. DESHALB DIE
KERNFRAGE: MIT WELCHER
WAHRSCHEINLICHKEIT IST
DER GETESTETE SCHULDIG,
WENN DER DETEKTOR
AUS... SCHLÄGT?



Die Verlässlichkeitsangaben zum Lügendetektor liefern uns die bedingten Wahrscheinlichkeiten

$$P(D|S) = 0,92$$

$$P(D|U) = 0,02$$

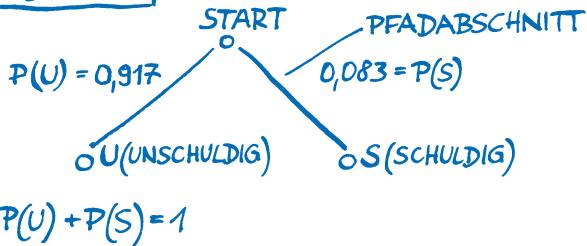
$P(D|S)$ ist die bedingte Wahrscheinlichkeit dafür, dass der Lügendetektor ausschlägt, wenn der Getestete schuldig ist. Analog ist $P(D|U)$ die bedingte Wahrscheinlichkeit dafür, dass der Lügendetektor bei einem Unschuldigen ausschlägt.

Wir können die Wahrscheinlichkeiten an den Pfaden eines Baumdiagramms antragen. Als **Pfad** bezeichnen wir einen Weg von der Wurzel über die Verzweigungspunkte des Baumdiagramms bis zur letzten Stufe.

Der Verzweigungspunkt, an dem das Baumdiagramm startet, ist die Wurzel des Baums. Von hier aus unterscheiden wir zwei Fälle: Eine Person ist schuldig oder unschuldig.

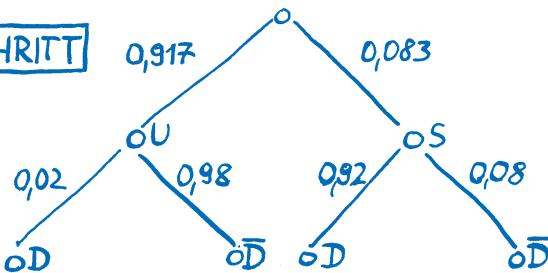
BAUMDIAGRAMM ZUR DARSTELLUNG DER BEDINGTEN WAHRSCHEINLICHKEITEN

1. SCHRITT



Jeden Pfadabschnitt versehen wir mit der Wahrscheinlichkeit, mit der das jeweilige Ereignis der nächsten Stufe eintritt. Die Summe der Wahrscheinlichkeiten aller von einem Verzweigungspunkt ausgehenden Pfadabschnitte muss 1 ergeben.

2. SCHRITT

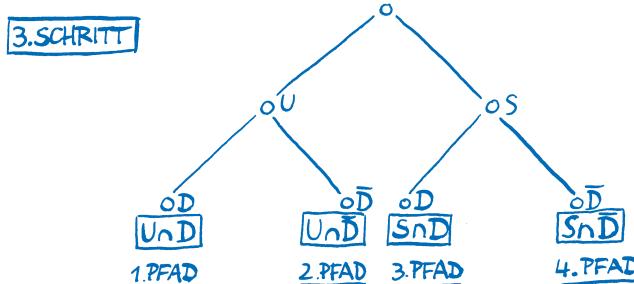


**DIE PFADABSCHNITTE IN DER ZWEITEN STUFE
WERDEN UNTERTEILT IN "AUSSCHLAG"(D) UND
"NICHT-AUSSCHLAG"(D-bar) DES LÜGENDETEKTORS.
ALLE VON EINEM VERZWEIGUNGSPUNKT AUSGEMEHDEN
ZWEIGE TRAGEN WAHRSCHEINLICHKEITEN,
DEREN SUMME „1“ IST.**

Auf der ersten Stufe unterscheiden wir die beiden Gegenereignisse *Schuldig* und *Unschuldig* mit 91,7% und 8,3%, was zusammen wieder 100% ergibt.

Die Wahrscheinlichkeit eines ganzen Pfads wird durch das Produkt seiner Pfadabschnitte berechnet. Die Wahrscheinlichkeit zum Beispiel, dass jemand unschuldig ist und der Lügendetektor trotzdem ausschlägt, ist das Produkt der beiden Wahrscheinlichkeiten

$$P(U) \cdot P(D|U) = 0,917 \cdot 0,02 \approx 0,018.$$



DIE AUSSAGE EINES PFADES GEWINNT MAN AUS DER „UND“-VERKNÜPFUNG ALLER EREIGNISSE, DIE AN DEN KNOTEN ABGETRAGEN SIND.

$U\&D$: PERSON IST UNSCHULDIG UND DETEKTOR SCHLÄGT AUS.

$U\~D$: PERSON IST UNSCHULDIG UND DETEKTOR SCHLÄGT NICHT AUS.

$S\&D$: PERSON IST SCHULDIG UND DETEKTOR SCHLÄGT AUS.

$S\~D$: PERSON IST SCHULDIG UND DETEKTOR SCHLÄGT NICHT AUS.

Wir haben genau zwei Pfade, in denen der **Lügendetektor ausschlägt**, das Ereignis D also eintritt:

Der Pfad, der bei $U \cap D$ mit der Wahrscheinlichkeit $P(U) \cdot P(D|U)$ endet, und der Pfad, der bei $S \cap D$ mit der Wahrscheinlichkeit $P(S) \cdot P(D|S)$ endet.

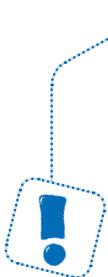
Zur Erinnerung: Wir suchen $P(S|D)$, die **bedingte Wahrscheinlichkeit**, mit der die getestete Person schuldig ist, wenn der Detektor ausschlägt. Das ist nicht zu verwechseln mit der bedingten Wahrscheinlichkeit $P(D|S)$, die eine Verlässlichkeitsaussage über den Detektor macht, wenn man weiß, dass der Getestete schuldig ist.

Die Wahrscheinlichkeit $P(S|D)$ können wir nicht direkt im Baumdiagramm ablesen. Es gibt keinen Pfad und auch keinen Pfadabschnitt, der ihren Wert repräsentiert.

Hier hilft die Formel von Bayes weiter.

FORMEL VON BAYES:

$$P(S|D) = \frac{P(S \cap D)}{P(U \cap D) + P(S \cap D)}$$



Die Idee dieser Formel ist die folgende: Es gibt mehrere „Ursachen“ für das Ausschlagen des Detektors (Ereignis D): schuldige und – fälschlicherweise – auch unschuldige Personen. Der **Nenner** enthält die Summe der Wahrscheinlichkeiten aller Pfade, in denen D zutrifft. Der **Zähler** enthält nur die Wahrscheinlichkeit eines Pfads, in dem D zutrifft, nämlich den, bei dem der Detektor bei einer schuldigen Person ausschlägt. Der Quotient berechnet also gerade den Anteil, den die schuldigen Personen am Ausschlag des Detektors haben.

**DER LÜGENDETEKTOR SCHLÄGT BEI 2
VON INSGESAMT 4 PFADEN AUS.**

$$1. \text{PFAD: } P(U \cap D) = P(D|U) \cdot P(U)$$

$$2. \text{PFAD: } P(S \cap D) = P(D|S) \cdot P(S)$$

EINGESETZT IN DIE FORMEL VON BAYES:

$$P(S|D) = \frac{P(S \cap D)}{P(U \cap D) + P(S \cap D)} = \frac{P(S) \cdot P(D|S)}{P(U) \cdot P(D|U) + P(S) \cdot P(D|S)}$$

Mit den obigen Werten ermitteln wir

$$P(S|D) = \frac{0,083 \cdot 0,92}{0,917 \cdot 0,02 + 0,083 \cdot 0,92} = 0,806 \approx 81\%$$

Das Ergebnis mag verblüffen: Obwohl der Lügendetektor ausschlägt, ist die getestete Person nur mit einer Wahrscheinlichkeit von rund 81% schuldig. Als Nachweis für eine Verurteilung reicht das wohl kaum aus.



Übrigens: Wer meint, der entscheidende Parameter, den man beim Lügendetektor zur Verbesserung der Aussage ändern müsste, sei die Verlässlichkeit, einen Schuldigen zu erkennen (momentan 92%), der irrt. Vielmehr muss man die Wahrscheinlichkeit, einen Unschuldigen irrtümlich als schuldig abzustempeln (momentan 2%), deutlich verkleinern. Woran liegt das?

Nun, die Anzahl der unschuldigen Personen ist viel größer als die Anzahl der schuldigen Personen. Diese Anzahlen gewichten aber in obiger Formel gerade die Verlässlichkeitswahrscheinlichkeiten. Somit schlägt sich die Unzuverlässigkeit von 2% ganz schön nieder.

Es kommt nicht nur auf die Verlässlichkeit eines Verfahrens, sondern auch auf die Größe der Population an. Ist die Population sehr groß (wie etwa die Zahl der Unschuldigen oder Gesunden in der Bevölkerung), so kann aus einer (scheinbar) hohen Verlässlichkeit bezogen auf das Individuum eine kleine Verlässlichkeit bezogen auf die ganze Population werden, was letztlich das Aus des Verfahrens (Lügendetektor, medizinischer Test) bedeuten kann.



Mit der Bayes'schen Formel können nicht nur Lügendetektoren qualifiziert werden. Für eine **medizinische Vorsorgeuntersuchung**, die beispielsweise auf Krebszellen oder auf HIV testet, ist eine Aussage, dass

eine getestete Person mit einer Verlässlichkeit von 81% erkrankt ist, völlig indiskutabel. Bei medizinischen Tests muss das Fehlerrisiko, einen Gesunden als krank einzustufen, so gering wie nur irgendmöglich gehalten werden.

Ein weiteres Anwendungsgebiet für die Formel von Bayes sind **Filter für das Aussortieren von Spam-E-Mails**. Das Vorkommen bestimmter Schlüsselwörter in einer E-Mail führt zu einer automatisierten Entscheidung darüber, ob die E-Mail erwünscht oder auszusortieren ist. Man nennt dieses Verfahren auch Bayes'scher Filter.



Während in Deutschland der Bundesgerichtshof 1998 den Einsatz von polygraphischen Untersuchungsmethoden („Lügendetektor“) im gerichtlichen Verfahren als Beweismittel generell ausgeschlossen hat, sind diese in den USA zugelassen. (So weit wir wissen, sind diese Untersuchungsmethoden auch in anderen EU-Ländern nicht zugelassen.) Selbst eine Reihe von schwerwiegenden Fehlentscheidungen, die sich auf die Verwendung von Polygraphen zurückführen ließen, konnten den Glauben an Lügendetektoren in den USA nicht erschüttern. So finden Lügendetektoren sogar beim Geheimdienst CIA und der Bundespolizei FBI Anwendung, um die Vertrauenswürdigkeit von Bewerbern zu beurteilen.



Thomas Bayes (1702 bis 1761) war ein englischer Mathematiker und presbyterianischer Pfarrer. Seine nach ihm benannte Formel, die er im Jahre 1750 entdeckte, wurde erst zwei Jahre nach seinem Tod veröffentlicht.

Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwortschutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: info@pearson.de

Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.**

Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

<http://ebooks.pearson.de>