Veröffentlicht in

Controller Magazin

2 / 2014

„Wahrscheinlichkeiten, Bayes-Theorem und statistische Analysen“

S. 68 – 74

Mit freundlicher Genehmigung der
Verlag für Controllingwissen VCW, Wörthsee-Etterschlag

www.vcw.de
www.haufe.de

Ein Service von: FutureValue Group AG eMail: Kontakt@FutureValue.de Internet: www.FutureValue.de

etwas näher eingegangen. Dabei werden auch Grenzen traditioneller statistischer Tests (und Alternativen) z.B. für die Ableitung von „Erfolgswahrscheinlichkeiten" erläutert.

Facetten des Wahrscheinlichkeits-Begriffs

Der Wahrscheinlichkeitsbegriff bezieht sich auf ungewisse Situationen, bei denen nicht mehr eindeutig auf genau ein bestimmtes Ergebnis geschlossen werden kann, sondern nur auf die Menge möglicher Ergebnisse (Ätialprinzip).

Nach der statistischen (frequentistischen) Auffassung wird die Wahrscheinlichkeit als empirisch gesicherte Aussage über die relative Häufigkeit des Eintretens eines bestimmten Ergebnisses verstanden, wobei eine ausreichend große Stichprobe und eine zeitstabile Wahrscheinlichkeitsverteilung unterstellt wird.[4]

Gemäß der subjektivistischen Wahrscheinlichkeitsauffassung drückt diese im Wesentlichen den Grad des Vertrauens einer Person („Entscheider") bezüglich der möglichen Realisierung eines bestimmten Ereignisses aus. Zwischen der frequentistischen und der subjektiven Auffassung des Wahrscheinlichkeitsbegriffs steht die Ableitung einer Wahrscheinlichkeit über das „Design".[5] Dabei wird die Wahrscheinlichkeit für das Eintreten eines Ereignisses abgeleitet aus bestimmten Basisinformationen (oder Annahmen) über den jeweiligen Sachverhalt. Wenn beispielsweise bekannt ist, dass mit einem „fairen" Würfel[6] gewürfelt wird, dessen sechs Seiten die Zahlen von „1" bis „6" aufweisen, kann die Wahrscheinlichkeit für das Auftreten einer „6" abgeleitet werden, ohne dass gewürfelt wird.

Im Beispiel beträgt die Wahrscheinlichkeit ein Sechstel. **Ähnlich lässt sich die Wahrscheinlichkeit einer Insolvenz (oder die Erfolgswahrscheinlichkeit einer Sanierungsstrategie) ableiten aus einem auf nachvollziehbaren Annahmen basierenden (strukturellen) Simulationsmodell des Unternehmens durch Anwendung der Monte-Carlo-Simulation.**[7]

Bei der traditionellen „frequentistischen" statistischen Auffassung wird die Annahme konstanter Verteilungsparameter (z.B. Standardabweichungen einer Verteilung) getroffen. Meist werden diese als bekannt vorausgesetzt.

Im Gegensatz zu diesen traditionellen Verfahren steht die **Bayes'sche Statistik. Hier wird keine Konstanz der Verteilungsparameter vorausgesetzt, sondern eine zeit- und zustandsabhängige Modellierung von Verteilungsparametern bzw. Risiken vorgenommen.** Die Parameter der die Risiken beschreibenden Wahrscheinlichkeitsverteilungen selbst werden somit als Zufallsvariablen aufgefasst.[8] Zudem können in der Bayes'schen Statistik sowohl subjektive Einschätzungen als auch objektive, gemessene Daten gleichzeitig berücksichtigt werden.[9] **In Anbetracht der häufig unvollständigen und unbefriedigenden historischen Daten, die die Verwendung subjektiver Schätzungen notwendig machen, erscheint der Bayes'sche Ansatz der Statistik als Grundlage für die Bewertung** (quantitative Beschreibung) **von Risiken besser geeignet**.[10] Der Bayes'sche Ansatz wird deshalb im Folgenden etwas näher erläutert, weil er einen nützlichen methodischen Rahmen bietet, um durch das Controlling ein „evidenzbasiertes" Management zu unterstützen.

Bedingte Wahrscheinlichkeiten und das Bayes-Theorem

Der Bayes'sche Ansatz unterscheidet zwischen den bereits vorliegenden (z.B. auch einfach subjektiv geschätzten oder aus dem „Design" abgeleiteten) Informationen über eine Wahrscheinlichkeitsverteilung (z.B. die Schätzung einer Eintrittswahrscheinlichkeit), den durch eine neu durchgeführte Erhebung oder ein Experiment neu beschafften Informationen und der sich aus der Verbindung dieser beiden ergebenden (neuen) A-posteriori-Verteilung.

Der Ansatz verdeutlicht und modelliert dabei einen optimalen Lernprozess, bei dem in Abhängigkeit des Vorliegens neuer Informationen (z.B. eines Experiments) die bestehenden Informationen einer Risikoquantifizierung verbessert werden. Der Bayes'sche Ansatz ist damit insbesondere geeignet, den in der Praxis der Unternehmen bei der Risikoquantifizierung erforderlichen Prozess des kontinuierlichen Lernens aufgrund neuer Informationen abzubilden.[11]

Mit Hilfe des so genannten Bayes-Theorems[12] lassen sich unbekannte Parameter einer Wahrscheinlichkeitsverteilung schätzen, Konfidenzintervalle für die unbekannten Parameter festlegen und Hypothesen über die Parameter prüfen. **Die Methodik des Bayes-Theorems, welche zur verbesserten Schätzung von Wahrscheinlichkeiten dienen kann, wird beispielsweise in der medizinischen Diagnose oder der Messung des Kreditausfallrisikos[13] angewendet**, während andere betriebswirtschaftliche Entscheidungen bisher selten mit dieser Methodik vorbereitet werden.[14]

Das Bayes-Theorem betrachtet so genannte **„bedingte Wahrscheinlichkeiten"**. Durch diese wird der Tatsache Rechnung getragen, dass im Allgemeinen eine Aussage davon abhängt, ob eine weitere Aussage wahr ist. Man schreibt A|B, um auszudrücken, dass A wahr ist unter der Bedingung, dass B wahr ist. Die Wahrscheinlichkeit von A|B, die bedingte Wahrscheinlichkeit, wird mit P(A|B) bezeichnet. Sie ist ein Maß für die Plausibilität oder Glaubwürdigkeit der Aussage A|B.

Grundsätzlich gibt das Bayes-Theorem an, wie die A-priori-Wahrscheinlichkeit einer Hypothese (Annahme) und bedingte Wahrscheinlichkeiten für bereits eingetretene Ereignisse zu kombinieren sind, um die A-posteriori-Wahrscheinlichkeit einer Hypothese zu bestimmen. Die A-priori-Wahrscheinlichkeit stellt dabei die Wahrscheinlichkeitsschätzung (über ein unbekanntes Ereignis) dar, die aufgrund von Vorwissen, z.B. über das „Design" (Erfahrungen), angenommen wird. Demgegenüber gibt die A-posteriori-Wahrscheinlichkeit die Wahrscheinlichkeit an, dass eine Hypothese nach bestimmten (neuen) beobachtbaren Ereignissen bestätigt wird. Mathematisch verhält sich die A-posteriori-Wahrscheinlichkeit, also P(A|B), für die Gültigkeit der Hypothese A nach dem Eintritt des Ereignisses B, wie folgt:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Anzahl Produkte …	Tatsächliche Häufigkeit	Test sagt „schadhaft“	Test sagt „OK“
mit Schaden	100	95	5
ohne Schaden	99.900	4995	94905
		Irrtumswahrscheinlichkeit	
		95/5090 = 1,87%	5/94905 = 0,01%

Abb. 1: Fallbeispiel

Zunächst ist die A-priori-Wahrscheinlichkeit für ein unbekanntes Ereignis gegeben (P(A)). **Dabei ist P(B|A) die Wahrscheinlichkeit für ein Ereignis B unter der Bedingung, dass A gilt – auch „Likelihood“ genannt – und P(A) ist die A-priori-Wahrscheinlichkeit für die Hypothese A.** Die A-priori-Wahrscheinlichkeit der Aussage über das unbekannte Ereignis wird also durch die Likelihood modifiziert, um die A-posteriori-Wahrscheinlichkeit zu erhalten. Aus Beobachtungen werden neue Informationen über dieses unbekannte Ereignis gewonnen, nämlich die bedingten Wahrscheinlichkeiten für ein Ereignis B unter der Bedingung, dass A gilt (P(B|A)). Daraus kann dann die A-posteriori-Wahrscheinlichkeit P(A|B) mit Hilfe des Bayes-Theorems bestimmt werden (Gleichung (1)). Sie drückt das Lernen aus (neuen) Erfahrungen aus.

Mit Hilfe des Bayes-Theorems ist es somit möglich, getroffene Wahrscheinlichkeitsannahmen, z.B. einer Risikoquantifizierung, mit neuen Daten zu korrigieren. Oder man kann Entscheidungen, die als richtig gelten, als falsch entlarven, denn Menschen neigen häufig dazu, ausschließlich neue Daten zu betrachten und die A-priori-Wahrscheinlichkeiten zu vernachlässigen. Es zeigen sich aber die Schwierigkeiten vieler Menschen, fundiert Wahrscheinlichkeiten und Likelihoods intuitiv zu bestimmen. Modelle und statistische Benchmarks können hier weiterhelfen.

Die Relevanz des Bayes-Theorems bei allen empirischen Tests (z.B. bei einer Risikoanalyse in Unternehmen oder medizinischen Diagnosen) oder statistischen Analysen wird deutlich, wenn man sich der beiden möglichen statistischen Fehler bewusst wird. **Ein statistischer Fehler erster Art liegt vor, wenn die formulierte Nullhypothese (Annahme) abgelehnt wird, obwohl diese wahr ist. Vom statistischen Fehler zweiter Art spricht man dagegen, wenn die Nullhypothese akzeptiert wird, obwohl diese falsch ist.** Bei Schlussfolgerungen aufgrund von empirischen Ergebnissen (z.B. eigenen Erfahrungen) wird oft die Wahrscheinlichkeit der sich aus diesen Erfahrungen (Datenhistorie) abgeleiteten Folgerungen überhaupt nicht betrachtet. Da es in der Realität kein endgültig gesichertes Wissen gibt, basieren Handlungen und Entscheidungen letztlich immer auf empirisch mehr oder weniger gut gesicherten Hypothesen, deren „empirische Bewährung“ (oder „Wahrscheinlichkeit“) natürlich offenkundig von Bedeutung ist (Poppers Falsifikationsprinzip im kritischen Rationalismus).

Bayes-Theorem und statistischer Test an einem Fallbeispiel

Wenn über Wahrscheinlichkeiten von Hypothesen, z.B. über Erfolgsfaktoren von Projekten, nachgedacht wird, wird meist nur über das so genannte Konfidenzniveau (Signifikanzniveau, SN) diskutiert.[15] **Das Signifikanz- oder Konfidenzniveau bezieht sich jedoch nur auf den Fehler erster Art.** Mit dieser Information alleine lassen sich jedoch keine sinnvollen Folgerungen ableiten – obwohl dies in empirischen ökonomischen Studien oder in den diagnostischen Tests der an sich sinnvoll evidenzbasierten Medizin regelmäßig geschieht.

Das Problem wird am folgenden Beispiel deutlich. Angenommen im Rahmen der Risikoanalyse eines Unternehmens wird ein **Testverfahren** eingesetzt, **das in 95% aller Fälle „korrekt“ einen zu erwartenden risikobedingten Schaden vorhersagt**, beispielsweise den Ausfall einer wichtigen Produktionseinheit oder einen Qualitätsmangel oder einen Fall von Untreue/Fraud.[16] **Eine 95%ige Sicherheit impliziert, dass tatsächlich 95% z.B. der Qualitätsmängel erkannt werden** (Signifikanzniveau), und zusätzlich wird angenommen, dass auch 95% der bezüglich der Qualität einwandfreien Produkte als solche erkannt werden.

Welche Schlussfolgerung lässt sich nun im Rahmen von Risikomanagement oder Qualitätsmanagement ziehen, wenn der Test positiv ist, also einen Fehler signalisiert. **Die intuitive Antwort lautet, dass mit 95%iger Sicherheit das betrachtete Teil fehlerhaft ist, weil der Test ja zu 95% „richtig liegt“. Bedauerlicherweise ist diese intuitive Einschätzung falsch.** Abbildung 1 verdeutlicht die Zusammenhänge:

Nehmen wir an, im Rahmen der Risikoanalyse (z.B. „Produkthaftpflicht“) werden in jedem Jahr 100.000 Produkte untersucht. Aus statistischen Daten der Vergangenheit ist nun bekannt, dass im Mittel 0,1% fehlerhaft sind. Durch den gewählten Test werden 95% der fehlerhaften Teile, also 95% x 100.000 x 0,001 = 95 auch als solche erkannt und 5 werden übersehen. Allerdings ist zu beachten, dass der Test auch in 5% aller Fälle ein an sich korrektes Produkt als fehlerhaft anzeigt – also einen Fehler zweiter Art vorliegt. Bei 100.000 untersuchten Teilen entspricht dies (100.000 – 100) x 5% also 4995 fälschliche Fehlermeldungen. Damit werden insgesamt 95 + 4995 = 5090 Teile als schadhaft ausgewiesen. Man erkennt unmittelbar, dass bei dem 95% sicheren Testverfahren lediglich 1,87% der Produkte, bei denen der Test einen Qualitätsfehler anzeigt, tatsächlich schadhaft sind. Für die quantitative Ri-

Autor

Dr. Werner Gleißner

ist Vorstand bei der FutureValue Group AG in Leinfelden-Echderdingen. Des Weiteren ist er im Vorstand der Risk Management Association e.V. in München.

E-Mail: kontakt@futurevalue.de

sikoanalyse und die Entscheidung zum Umgang mit den als „fehlerhaft" klassifizierten Teilen muss dies berücksichtigt werden. Zudem muss entschieden werden, ob ein Testverfahren mit dieser Qualität überhaupt ökonomisch nützliche Ergebnisse liefert, was im Beispiel von der möglichen Schadenshöhe im Haftpflichtfall abhängt.

Das hier abgeleitete Rechenergebnis des Fallbeispiels – und das Qualitätsproblem des Tests – lässt sich auch herleiten bei unmittelbarer Verwendung des Bayes-Theorems gemäß Formel (1). Mit obigen Angaben gilt das Ereignis B („Test benennt Produkt als schadhaft") und Ereignis A („Produkt ist schadhaft") hier P(A)=0,1% und P(B|A)=95%. P(B) zeigt die Wahrscheinlichkeit, dass der Test „schadhaft" angibt (95+4995)/100.000 =5,09% (vgl. Abbildung 1).

(2) $P(A \mid B) = \frac{0{,}95 \cdot 0{,}001}{0{,}0509} = 0{,}0187 = 1{,}87\%$

Wie die Rechnung mit dem Bayes-Theorem zeigt, liegt die Wahrscheinlichkeit P(A|B) bei den schon berechneten 1,9%.

Mit Hilfe der Bayes-Formel wird also eine geschätzte Eintrittswahrscheinlichkeit eines Ereignisses modifiziert in Abhängigkeit des Vorliegens zusätzlicher Indizien (Erfahrungen, Forschungsergebnisse). Sie ist von zentraler Bedeutung für die Schätzung von Wahrscheinlichkeiten im Rahmen des Risikomanagements z.B. bei der Quantifizierung der Eintrittswahrscheinlichkeit versicherter Risiken[17] – und allgemein für das Lernen aus Erfahrungen.

Aussagekraft statistischer Tests und Erfolgsprognosen

Für eine Beurteilung des – positiven oder negativen – Testergebnisses ist es also grundsätzlich erforderlich, drei Informationen zu besitzen, nämlich

- das Signifikanzniveau (SN, das bekannte „P" der statistischen Tests), hier also ca. 5%, das den Umfang des Fehlers erster Art angibt **P(B)**,
- die „Power" des Tests, die die Möglichkeit der Vermeidung des Fehlers zweiter Art angibt, hier also 95% **P(B|A)** und

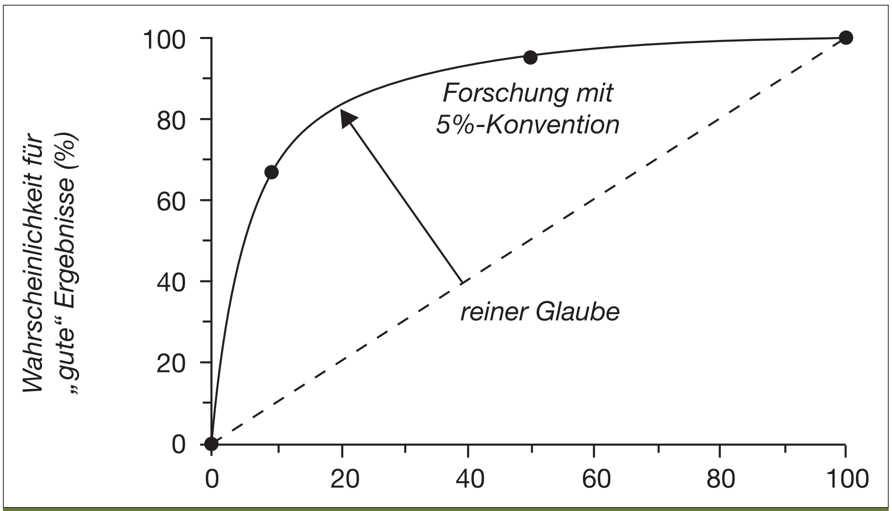

Abb. 2: Die Wahrscheinlichkeit „gute" Ergebnisse hängt von der Wahrscheinlichkeit für gute Ideen ab. Diese Abbildung gilt bei einem Signifikanzniveau von 0,05 (fünf Prozent).[22]

- die Inzidenz, also die Häufigkeit des gesuchten Merkmals („Schäden") in der Grundgesamtheit **(P(A))**.

Tatsächlich werden jedoch bei empirischen Untersuchungen und diagnostischen Tests – wie erwähnt – meist lediglich Signifikanz- bzw. Konfidenzniveau betrachtet, und schon die Betrachtung der „Power" ist selbst in wissenschaftlichen Untersuchungen noch nicht selbstverständlich. Noch gravierender ist, dass die Inzidenz in der Regel unbekannt ist, was die Anwendung derartiger statistischer Tests – insbesondere der üblichen Signifikanztests z.B. bei Regressionsanalysen – aus Sicht mancher Kritiker erheblich in Frage stellt.[18]

Trotz der Schwächen kann der durchgeführte Test zur Produktqualität im vorherigen Fallbeispiel jedoch dazu beitragen, die A-priori-Wahrscheinlichkeit des Auftretens eines Fehlers (hier 0,1%) zu verändern. Diese Modifikation einer A-priori-Wahrscheinlichkeit durch eine Zusatzinformation, dem Testergebnis bzw. einer neuen Erfahrung, wird gerade durch das Bayes'sche Theorem verdeutlicht. Wie erwähnt, beträgt die Wahrscheinlichkeit eines Schadens nach dem Test (a posteriori) 1,87%. Den Informationsgewinn durch einen Test kann man durch das so genannte Information-Ratio (IR) quantifizieren.[19] Das Information-Ratio[20] (Likelihood Ratio) ist gerade der Quotient aus „Power" und dem Signifikanzniveau (SN) des Tests, im Beispiel also

(3) $IR = \frac{Power}{SN} = \frac{P(B \mid A)}{P(B)} = \frac{0{,}95}{0{,}0509} = 18{,}7$

Mit Hilfe des Information-Ratios (IR) eines Tests kann man die „neue" Wahrscheinlichkeit (nach Durchführung des Tests) im Vergleich zur Situation vorher wie folgt beschreiben (vgl. Gleichung (1)):

bzw.

(4) $P(A \mid B) = P(A) \cdot IR = 0{,}1\% \cdot 18{,}7 = 1{,}87\%$

Zur Interpretation des Resultats eines empirischen Tests ist es also erforderlich, eine Vorabinformation (ein „Vorurteil") über die grundlegende Wahrscheinlichkeit zu haben. Bei empirischen wissenschaftlichen Tests besteht hier das grundlegende Problem, dass die „Grundwahrscheinlichkeit" (Inzidenz) z.B. „guter Ideen", deren Nutzen empirisch getestet wird, nicht bekannt ist. Beck-Bornholdt und Dubben (2007) führen dazu aus:

„Wenn ich nur untaugliche Ideen habe, dann ist die Wahrscheinlichkeit für ein echtes Ergebnis gleich Null. Wenn ich als brillante Forscherin nur taugliche Ideen habe, ist die Wahrscheinlichkeit für ein echtes Ergebnis entsprechend 100%." [21]

So kann beispielsweise bei medizinischen Untersuchungen zur Wirksamkeit von Medikamenten aus einem „signifikant positiven" Testergebnis nicht abgeleitet werden, mit welcher Wahrscheinlichkeit das Medikament tatsächlich eine (bessere) Wirkung zeigt. Es ist nämlich schlicht nicht bekannt, wie hoch der Prozentsatz aller (besser) wirksamen Medikamente an der Grundgesamtheit aller (testbaren) Medikamente ist – also ist die Quote der „guten Ideen" unbekannt. **Die „statistische Signifikanz"**

	San-A 90% Erfolgsrate		San-B 10% Erfolgsrate	
Wohin schickt Herr Meier das erste Krisenunternehmen: (Münzwurf)?	50%		50%	
Ergebnis für das erste Krisenunternehmen	Erfolg 45%	Misserfolg 5%	Misserfolg 45%	Erfolg 5%
Konsequenz	Zweiter Kunde zu San-A	Zweiter Kunde zu San-B	Zweiter Kunde zu San-A	Zweiter Kunde zu San-B
Wohin schickt Herr Meier das zweite Krisenunternehmen?	90%		10%	
Ergebnis für das zweite Krisenunternehmen	Erfolg 81%	Misserfolg 9%	Misserfolg 9%	Erfolg 1%
Konsequenz	Dritter Kunde zu San-A	Dritter Kunde zu San-B	Dritter Kunde zu San-A	Dritter Kunde zu San-B
Wohin schickt Herr Meier das dritte Krisenunternehmen?	90%		10%	

Abb. 3: Algorithmus im Einsatz als Never-change-a-winning-team-Strategie

drückt also nicht das aus, was meist hineininterpretiert wird. Statistische Signifikanz drückt Folgendes aus:

„Wie wahrscheinlich ist das beobachtete Ergebnis, gesetzt den Fall, das untersuchte Medikament hat keine Wirkung.“

Diese Aussage ist klar zu unterscheiden von der Aussage:

„Wie wahrscheinlich ist es, dass das Medikament Wirkung hat.“

Zwischenfazit: **Intuitive Bauchentscheidungen versagen oft, was viele weitere psychologische Untersuchungen zeigen**[23] – und Entscheidungen basierend auf empirischen Untersuchungen bedürfen sehr viel Sorgfalt. **Notwendig ist eine Vorabinformation, ein Prior, z.B. ein Branchenbenchmark, der mit zusätzlichen Erkenntnissen (Erfahrungen) aus dem eigenen Unternehmen schrittweise verbessert wird.**

Eine mögliche Reaktion bei empirischen Untersuchungen, mit deren Hilfe man Gesetzmäßigkeiten (z.B. für unternehmerische Entscheidung oder auch die medizinischen Behandlungen) ableiten möchte, kann darin bestehen – als Faustregel – im Rahmen der Studien sehr hohe Signifikanzniveaus (von z.B. 0,1%) bei gleichzeitig sehr hoher Power des Tests (von 90%) zu fordern.[24, 25] In der Unternehmenspraxis ist zu empfehlen, diese Annahmen zu diskutieren und möglichst Konsens herzustellen (z.B. bezüglich Erfolgswahrscheinlichkeit bestimmter Projekttypen).

Alternativen zu statistischen Tests: Lernen aus Erfahrungen

Schon mit sehr einfachen (auch subjektiv geschätzten) „Vorinformationen“ kann man die Qualität von Entscheidungen verbessern.[26] Wenn beispielsweise zwei Entscheidungsoptionen zur Verfügung stehen, bezüglich derer nacheinander zu entscheiden ist, lässt sich das folgende Verfahren anwenden: Angenommen es ist zu entscheiden, ob bei einem Händler A (zum Preis P_A) ein bestimmtes Produkt gekauft werden soll. Bei einer Entscheidung für Händler A ist damit grundsätzlich der Preis P_A fällig. Entscheidet man sich gegen Händler A, ist ein Kauf bei Händler B zum Preis P_B notwendig. Angenommen wird nun, dass bei der ersten Entscheidung – für oder gegen Händler A – der Preis des Händlers B noch nicht bekannt ist. Man kann sich beispielsweise vorstellen, dass man den Preis für Benzin an einer Autobahn-Tankstelle sieht, aber noch nicht weiß, welcher Preis bei der nächsten Tankstelle zu bezahlen wäre. Soll man sich nun für den Händler A entscheiden? Ohne jegliche Vorkenntnis steht die Wahrscheinlichkeit, die bessere Entscheidung zu treffen bei 50%. Bei einer Vorinformation, also einer Hypothese über den „angemessenen Preis" (P_C) kann man diese Chance verbessern. Es gilt die einfache Regel: Man schätzt vorab einen angemessenen Preis P_C. Bietet Händler A einen niedrigeren Preis, kauft man bei ihm. Andernfalls wird dessen Angebot abgelehnt und entsprechend bei Händler B gekauft.

Als Alternative zur Anwendung statistischer Tests speziell bei Unkenntnis über deren „Power" und „Inzidenz" („Benchmark") schlagen Beck-Bornholdt und Dubben (2007) eine einfache „Faustregel" vor.[27] Sie empfehlen eine „Never-change-a-winning-team-Strategie" mit einem ergänzenden Lernalgorithmus. Folgendes Beispiel verdeutlicht die Idee: Die Problemstellung besteht darin, als Sanierungsspezialist einer Bank (Herr Meier) ohne Vorkenntnis zu entscheiden, welchen Sanierungsberater man seinen insolvenzbedrohten Firmenkunden empfehlen soll.

Ähnlich ist die Situation, wenn z.B. über eine Werbeagentur, einen internen Projektleiter oder eine Verkaufsförderungsmaßname zu entscheiden ist. Das Kundenunternehmen im Fallbeispiel befindet sich aufgrund schlechter wirtschaftlicher Rahmenbedingungen in einer bestandsgefährdenden Situation. Es wird angenommen, dass Herr Meier über die Erfolgswahrscheinlichkeit der Sanierungsberater (San-A und San-B) zunächst nichts weiß, also beide zunächst identisch mit z.B. jeweils 50% einschätzen muss. Die Erfolgsraten beider Sanierungsberater unterscheiden sich tatsächlich jedoch deutlich, 90% vs. 10% Erfolgswahrscheinlichkeit im Durchschnitt der ähnlichen Sanierungsfälle. Als Erfolg wird dabei gewertet, wenn das Unternehmen nach einem Jahr noch existiert und mindestens ein „B-Rating" erreicht. Mit zunehmender Zahl von Erfahrungen aus Sanierungsfällen lässt sich die Wahrscheinlichkeit des Erfolgs eines Sanierungsberaters abschätzen (siehe Abbildung 3); diese Abbildung[28] verdeutlich den Ablauf.

Für das zweite sanierungsbedürftige Unternehmen verfährt Herr Meier nun nach einer einfachen Regel: Im Falle des Erfolgs des ersten Sanierungsberaters wird das zweite Krisenunternehmen zum gleichen Sanierungsberater empfohlen. Im Falle des Misserfolgs an den jeweils anderen Sanierungsberater. Die Wahrscheinlichkeit für einen Erfolg für das zweite Unternehmen nach dieser Entscheidungsregel liegt bereits bei 81%, also durchaus in der Nähe der Erfolgsrate des besseren der beiden Sanierungsberater.[29]

Sollte er den ersten Sanierungsfall an San-B empfehlen, dem weniger leistungsfähigen Berater, wird dies mit überwiegender Wahrscheinlichkeit ein Fehlschlag, sodass die nächste Empfehlung San-A sein wird. Insgesamt zeigt sich, dass mit einem vergleichsweise einfachen (und plausiblen) Verfahren durchaus eine interessante Alternative zu statistischen Tests besteht. Die „Never-change-a-winning-team"-Faustregel ist robust und insbesondere auch anwendbar, wenn keine historischen Daten vorliegen oder die „Inzidenz"[30] von „Erfolg" (in der Grundgesamtheit) nicht bekannt ist.

Die einfach anwendbaren Faustregeln wie „Never-change-a-winning-team" und „Take-the-best" (Gigerenzer (2005)) verdeutlichen, wie mit einfachen und robusten Regeln Entscheidungen unterstützt werden können. Anzumerken ist hier immer, dass idealerweise der Einsatz derartiger Faustregeln zusätzlich abgesichert werden kann durch vertiefende (aber wie hier erläutert auch nicht zwingend bessere) wissenschaftlich-statistische Methoden.

Fazit und Praxisempfehlung

Zusammenfassend ist festzuhalten, dass die Bayes-Statistik eine geeignete Grundlage zur Beantwortung vieler quantitativer Fragestellungen z.B. in Controlling und Risikomanagement darstellt. In der Bayes-Statistik wird die Vorläufigkeit jeden Wissens, speziell auch jeder aus statistischen Daten der Vergangenheit abgeleiteten Wahrscheinlichkeitsschätzung, betont. **Die Quantifizierung von Risiken wird als Lernprozess aufgefasst**, bei dem aufgrund neuer Erkenntnisse Wahrscheinlichkeitsschätzungen angepasst werden. Aufgrund der erheblichen Schwierigkeiten von Menschen, das zugrundeliegende Bayes-Theorem intuitiv korrekt anzuwenden, sollte ein statistisches Lernen immer durch eine explizite Anwendung des Theorems erfolgen. Aber auch (explizit formulierte) Lern-Heuristiken, wie z.B. „Never-change-a-winning-team", können nützlich sein.

Bei der Entscheidungsvorbereitung für Vorstand bzw. Geschäftsführung durch das Controlling ist Transparenz über getroffene Annahmen, z.B. Erfolgs- oder Schadenswahrscheinlichkeiten, zwingend notwendig. Empfehlenswert ist ergänzend, diese Annahmen möglichst gut abzusichern, durch Benchmarks und statistische Analysen, weil Menschen – wie die psychologische Forschung zeigt – Daten und Erfahrungswerte der Vergangenheit nicht gut „intuitiv" auswerten können („evidenzbasiertes Management"). Auch bei der Anwendung statistischer Tests und (Regressions-) Analysen gibt es jedoch Fallen, die man kennen sollte; aber ebenso existieren ergänzend nutzbare einfache Methoden, um aus Erfahrungen der Vergangenheit systematisch (!) zu lernen.

Literaturverzeichnis

Alexander, C. (2003): Operational Risk – Regulation, Analysis and Management, London 2003.

Beck-Bornholdt, H.-P./Dubben, H.-H. (2007): Der Schein der Weisen, 5. Auflage, Rowohlt Taschenbuch Verlag, Reinbek 2007.

Brückner, R./Gleißner, W. (2013): Unbefriedigende Datenlage: Ein Argument für den Ausbau von Controlling- und Risikomanagement-Methoden, in: Controller Magazin, Juli/August, Ausgabe 4, S. 12-16.

Buzzel, R. D. / Gale, B. T. (1989): Das PIMS-Programm: Strategien und Unternehmenserfolg, Gabler Verlag, Wiesbaden 1989.

Gigerenzer, G. (2005): I think, therefore I err, in: Social Research, 1/2005.

Gigerenzer, G./Kober, H. (2008): Bauchentscheidungen, Die Intelligenz des Unbewussten und die Macht der Intuition, 1. Auflage, Goldmann Verlag.

Gleißner, W. (2009): Metarisiken in der Praxis: Parameter- und Modellrisiken in Risikoquantifizierungsmodellen, in: RISIKO MANAGER, 20/2009, S. 14-22.

Gleißner, W. (2011): Grundlagen des Risikomanagements im Unternehmen, 2. Auflage. Vahlen.

Gleißner, W. (2013a): Die risikogerechte Bewertung alternativer Unternehmensstrategien: ein Fallbeispiel jenseits CAPM, in: Bewertungspraktiker, 3 / 2013, S. 82-89.

Gleißner, W. (2013b): Die Sanierungserfolgswahrscheinlichkeit im IDW S 6 aus betriebswirtschaftlichentscheidungsorientierter Perspektive, in: KSI – Krisen-, Sanierungs- und Insolvenzberatung, 4 / 2013, S. 172-174.

Gleißner, W. / Nguyen, T. (2013): Prämienkalkulation und die „Sanierung" von Versicherungsverträgen, in: Zeitschrift für die gesamte Versicherungswissenschaft, Oktober 2013, DOI 10.1007/s12297-013-0248-0.

Koch, K. R. (2000): Einführung in die Bayes-Statistik, Springer Verlag.

Liekweg, A. (2003): Risikomanagement und Rationalität, Gabler Verlag, Wiesbaden 2003.

Metzler von, L. (2004): Risikoaggregation im industriellen Controlling, EUL Verlag, Lohmar.

Newman, J./Pearson, E. (1933): On the problem of the most efficient tests of statistical hy-potheses, in: Philosophical Transactions of the Royal Society of London A, 231/1933, S. 289-337.

Norland, E./Stabile, D. (2000): Bayesian approaches to finance, in: Global Investor, September 2000.

Pfeffer, J./Sutton, R. I. (2007): Harte Fakten – Gefährliche Halbwahrheiten und absoluter Unsinn, Addison-Wesley Verlag, München, 2007.

Rieder, M. (2005): Bayesianisches Kredit-Scoring zur Messung des Ausfallrisikos, in: Romeike, F. (Hrsg.), Modernes Risikomanagement: Die Markt-, Kredit- und operationellen Risiken zukunftsorientiert steuern, WILEY VCH Verlag, Weinheim, 2005.

Sinn, H. W. (1980): Ökonomische Entscheidungen bei Ungewissheit, J.C.B. Mohr (Paul Siebeck), Tübingen

Slovic, P. (2004): What's fear got to do with it? It's affect we need to worry about, in: Missouri Law Review, 69, S. 971-990.

Sterne, J.A./Smith, G.D. (2001): Sifting the evidence – what's wrong with significance tests?, in: British Medical Journal, 322/2001, S. 226-231.

Taleb, N. N. (2008): Der Schwarze Schwan: Die Macht höchst unwahrscheinlicher Ereignisse, Carl Hanser Verlag, München 2008.

Tversky, A./Kahneman, D. (1979): Prospect Theory: An Analysis of Decision under Risk, in: Econometrica, Vol. 47, No. 2. 1979, S. 263-292.

Zelen, M. (1969): Play-the-winner rule and the controlled clinical trial, in: Journal of the American Statistical Association, 64/1969, S. 131-146.

Zhou, S. (2005): Auf dem falschen Dampfer? Anwendung des Bayes-Theorems zur Risiko-Abschätzung in der strategischen Unternehmensplanung, in: RISKNEWS, 06/2005.

Fußnoten

[1] Siehe Gleißner (2011), S. 139f.

[2] Siehe Gleißner (2013a) und Gleißner (2013b).

[3] vgl. Buzzel/Gale (1989); Pfeffer/Sutton (2007) und Zhou (2005).

[4] Zum Problem der Annahme: Taleb (2008).

[5] Siehe Gigerenzer/Kober (2013).

[6] z.B. mit homogener Dichte

[7] In der Praxis werden aber wiederum einige der zugrundeliegenden Annahmen nur plausibel sein, so dass hier ein gradueller Übergang zu subjektiven Wahrscheinlichkeitskonzepten besteht.

[8] Grundsätzlich besteht auch die Möglichkeit, eine Unsicherheit bezüglich einer geschätzten Wahrscheinlichkeit im Entscheidungskalkül zu berücksichtigen. Siehe Sinn (1980); Gleißner (2009), S. 14-22 sowie Liekweg (2003).

[9] Siehe Norland/Stabile (2000), S. 4-5; Koch (2000), sowie von Metzler (2004), S. 69.

[10] Zum Umgang mit unbefriedigenden Daten bei der Risikoquantifizierung siehe Brückner/Gleißner (2013).

[11] Siehe weiterführend Alexander (2003), S. 129-170.

[12] Vgl. Zhou (2005), S. 71-74.

[13] Vgl. Rieder (2005), S. 185-200.

[14] Siehe Zhou (2005), S. 71.

[15] Das Signifikanzniveau wird – irreführenderweise – oft auch als „Irrtumswahrscheinlichkeit" bezeichnet, siehe hierzu Beck-Bornholdt/Dubben (2007), S. 184.

[16] Z.B. basierend auf statistischen Datenanalysen z.B. unter Nutzung des Bendford'schen Gesetzes.

[17] Vgl. Gleißner, W./Nguyen, T. (2013).

[18] Siehe Beck-Bornholdt/Dubben (2007), S. 153-204.

[19] Siehe Beck-Bornhold/Dubben (2007), S. 227-236.

[20] Bei Labortests gilt, dass der Informationsertrag (Likelihood Ratio) gerade dem Quotienten aus Sensitivität und (1 – Spezifität) entspricht.

[21] Quelle: Beck-Bornholdt/Dubben (2007), S. 172-174.

[22] Quelle: Beck-Bornholdt/Dubben (2007), S. 173.

[23] Vgl. z.B. Slovic (2004), Tversky/Kahnemann (1979).

[24] Siehe Sterne und Smith (2001).

[25] Siehe Beck-Bornholdt/Dubben (2007), S. 213-222.

[26] In Anlehnung an Beck-Bornholdt/Dubben (2007), S. 293-240.

[27] In Anlehung an Beck-Bornholdt/Dubben (2007), S. 241-254.

[28] Aus Beck-Bornholdt/Dubben (2007), S. 246.

[29] Die Erfolgsrate konvergiert langfristig gegen 82% (vgl. Beck-Bornholdt/Dubbe (2007)).

[30] Bedeutet in der medizinischen Statistik die Anzahl der Neuerkrankungen in einer Bevölkerungsgruppe an einer bestimmten Krankheit während einer bestimmten Zeitspanne.