Risiko – ein Überlebensratgeber

MANFRED BOROVCNIK, KLAGENFURT

Risiko ist allgegenwärtig. Wie wird Risiko wahrgenommen? Wir diskutieren aktuelle Beispiel zu Risiko und werden besonders Situationen aus dem Gesundheitsbereich aufgreifen. Hier gilt, wie auch in anderen Zusammenhängen, dass mehr Wissen und mehr Daten zu größerer Unsicherheit führen können. Das unterscheidet Risiko vom verwandten Konzept der Wahrscheinlichkeit. Es spielen natürlich die verschiedenen Ansätze, Wahrscheinlichkeit zu interpretieren eine Rolle; genauso wichtig ist, zu unterscheiden, wer von der Modellierung betroffen ist. Entscheidungen etwa, die für ein System rational sind, müssen für einzelne Personen durchaus nicht zielführend sein. Kann es einen Ratgeber, mit Risiko umzugehen, geben?

1. Risiko

Unter Risiko verstehen wir eine Situation mit inhärenter Unsicherheit über (zukünftige) Ausgänge, welche mit einem Impakt (Kosten, Schaden, Vorteil) verbunden sind. Manchmal wird zum Vergleich mehrerer Entscheidungsoptionen der erwartete Wert genommen. Wir werden aber auch andere Kriterien ansprechen. Für eine Definition von Risiko und einen Versuch der Abgrenzung gegenüber verwandten Begriffen wie Hazard verweisen wir auf Borovcnik (2015). Risiko wird uneinheitlich verwendet; neben vielen Varianten können wir folgende Extremfälle unterscheiden:

- Risiko kann sich lediglich auf die Wahrscheinlichkeit des unangenehmen Ereignisses beziehen, ohne dass man die Folgen (Kosten) mitbedenkt.
- Risiko kann sich auch nur in Bezug auf die Folgen einer unsicheren Situation beziehen, ohne dass man die damit verbundenen Wahrscheinlichkeiten einbezieht.

Je nachdem, wer in einer Situation mit Risiko beteiligt ist, kann sich die Wahrnehmung und Beurteilung von Risiken völlig anders darstellen. Eine zu treffende Entscheidung mag sich unterschiedlich auswirken (Borovcnik & Kapadia, 2011b):

- Die Entscheidung kann eine Person allein betreffen. Das ist insofern einfach, als dieselbe Person, welche die Entscheidung trifft, auch von den Konsequenzen betroffen ist. Eine Person mag das Risiko einer gefährlichen Bergtour auf sich nehmen, weil die Freude daraus die Risiken überwiegt. Allerdings, wenn ein Unfall passiert, sind auch andere Personen mitbeteiligt; etwa werden die Kosten der Behandlung teilweise von der Krankenversicherung, also der Öffentlichkeit getragen.
- Die Entscheidung mag zwischen zwei oder mehreren Stakeholdern geteilt werden. In der Frage, ob ein Screening-System zur Früherkennung einer bestimmten Krankheit eingeführt und beworben werden soll, treffen sich verschiedene Interessensgruppen: die Gesamtheit der Menschen, welche diese Krankheit erleiden könnten, die Ärzteschaft, das Gesundheitswesen, Institutionen, die mit der Finanzierung öffentlicher Einrichtungen betraut sind, u.a.m.
- Die Entscheidung mag zwischen (einzelnen) Menschen und einer Institution geteilt werden. Wenn sich eine Person für das Screening auf eine bestimmte Krankheit entscheidet, trifft sie auf das Gesundheitssystem als Ganzes. Es mag sein, dass für das Gesundheitssystem als übergeordnete Einheit eine Entscheidung für das Screening vernünftig ist und dennoch für die einzelnen Personen – ja sogar für alle einzelnen Personen – eine Entscheidung pro Screening schlecht ist.

Speziell wenn ein Einzelner auf eine Institution trifft, sind die Folgen völlig unterschiedlich. Ein Arzt als Vertreter des Gesundheitssystems hat eine andere Art von Verantwortlichkeit (auch gesetzliche Haftpflicht) als die Einzelperson. Der Arzt kann daher bestimmte Vorgehensweisen, die für den Einzelnen durchaus als Option offen stehen, gar nicht einmal andenken. Der Arzt hat darüber hinaus auch Interessen (natürlich auch finanzieller Art; aber auch Einfluss zu halten oder zu verstärken, etc.). Andererseits hat der Patient die Folgen der Entscheidung ganz persönlich zu tragen.

Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), Heft 49, 2016, S. 1-16.

1.1 Versuche, Risiko zu definieren

Die Begrifflichkeiten, die mit Risiko in Beziehung stehen, sind derart unterschiedlich festgelegt, dass man eher nur von Versuchen sprechen kann, Risiko zu definieren, wenn man mit der Festlegung auch den Gebrauch des Begriffes in verschiedenen Kontexten einschließen will. Wir orientieren uns an Borovcnik (2015):

1) Ein unerwünschtes Ereignis, das eintreten oder ausbleiben kann.
2) Die Ursache eines unerwünschten Ereignisses, das eintreten oder ausbleiben kann.
3) Die Wahrscheinlichkeit eines unerwünschten Ereignisses, das eintreten oder ausbleiben kann.
4) Der erwartete Wert eines unerwünschten Ereignisses, das eintreten ausbleiben kann.
5) Der Umstand, dass eine Entscheidung unter Bedingungen bekannter (und nicht unbekannter) Wahrscheinlichkeiten getroffen wird.

Die Wahrnehmung von Risiken hängt eng mit der Einschätzung von Wahrscheinlichkeiten zusammen. So etwa ist bekannt, dass gemeinsames Auftreten von hohen Folgen und kleiner Wahrscheinlichkeit die Wahrnehmung und Beurteilung dieser Wahrscheinlichkeit verzerrt.

An Verwandten von Risiko sind folgende Begriffe zu nennen:

- Risiko und Hazard: Hazard ist dabei „Ursache“ für ein unerwünschtes Ereignis.
- Risk und Uncertainty: Knight (1921) sieht den Unterschied zwischen Risk und Uncertainty darin, ob Wahrscheinlichkeiten bekannt sind. Für ihn sind Risiken mit bekannten Wahrscheinlichkeiten verbunden und damit berechenbar in dem Sinne, dass man bestimmte Größen ausrechnen kann, die für Entscheidungen wichtig sind.
- Risiko und Nutzen in ökonomischer Theorie: Zwei Dinge sind hier besonders erwähnenswert: Zum einen das Prinzip, den erwarteten Nutzen zu maximieren, das in vielen Fällen als ein Entscheidungskriterium fungieren kann, weil dadurch die eine oder andere Entscheidung als besser eingestuft wird. Zum anderen die Beschreibung von Verhalten von Personen aus ihrer Nutzenfunktion u; etwa wird das als Risiko-Aversion bekannte Verhalten durch u''/u' festgehalten.
- Systemanalytischer Ansatz zur Untersuchung von Situationen mit Risiko: Der Ansatz zur Erfassung von Risiko umfasst die folgenden Komponenten bzw. Fragen: Welche Konstituenten machen Risiko aus? Wer trägt das Risiko, wer ist involviert? Welchen Charakter hat das Risiko: etwa eine nicht erwünschte Nebenwirkung von Handlungen, die wir unternehmen müssen, um anderweitigen Schaden abzuwehren (etwa eine Operation nach einem Unfall); oder eine dräuende Gefahr, nicht real, aber vorstellbar (etwa eine mögliche Krebserkrankung der Cervix, wenn die Frau nicht rechtzeitig eine HPV-Impfung erhält). Welche Information hat man zur Beurteilung der Risiken zur Verfügung? Als Situation kann man sich etwa die Entscheidung vorstellen, ob sich die Eltern eines zehnjährigen Mädchens für oder gegen eine HPV-Impfung entscheiden. Ist es eine sinnvolle Maßnahme zur Vorsorge oder eine durch drängende und drastische Information der möglichen negativen Folgewirkungen seitens des Gesundheitswesens erzwungene Entscheidung? (Siehe auch Borovcnik & Kapadia, 2011a; für Brustkrebs, siehe Borovcnik, 2016).

1.2 Risiko in der Statistik – Risiko von Fehlentscheidungen

In der beurteilenden Statistik werden üblicherweise die Fehlentscheidungen beim Testen von Hypothesen als Risiko bezeichnet. Dabei allerdings bezieht sich der Begriff Risiko einerseits allein auf den Typ der Fehlentscheidung, andererseits auf die Wahrscheinlichkeit, einen solchen Fehler zu begehen.

Aufgabe 1 (Qualitätsregelung): Im Rahmen der Qualitätsregelung soll verhindert werden, dass zu viel Ausschuss durchschlupft. Bauteile können defekt (Ausschuss) oder nicht-defekt (kein Ausschuss) sein. Sagen wir, eine Ausschussquote von 4% sei vertretbar, während 10% Ausschuss zwar nicht sicher (das ist nicht möglich), aber mit sehr hoher Wahrscheinlichkeit erkannt werden soll. Wir wollen eine Entscheidungsregel für die stündliche Kontrolle angeben, die langfristig verhindert, dass die Produktion nicht zu oft unterbrochen wird, wenn die Ausschussquote 4% nicht überschreitet, und die zusätzlich garantiert, dass eine Abweichung der Qualität (es liegt aktuell schon 10% Ausschuss vor) relativ sicher erkannt wird.

Die Situation ist eigentlich ein statistisches Testproblem der Nullhypothese H_0: $p_0 = 0{,}04$ gegen die Alternativhypothese $p_1 = 0{,}10$. Wenn wir bei 100 untersuchten Bauteilen entscheiden, die Produktion bei $X \geq 8$, d.h., bei 8 und mehr defekten Teilen in der Stichprobe, zu unterbrechen, so können wir folgende Größen berechnen:

$$P(X \geq 8 \mid \mathrm{H}_0 : p = 0{,}04) = 0{,}0475 \text{ und } P(X \leq 7 \mid \mathrm{H}_1 : p = 0{,}10) = 0{,}2061.$$

Wir können feststellen, dass in rund 4,8% der Kontrollzeitpunkte, zu denen die Produktion eigentlich unter Kontrolle ist, die Fertigung gestoppt (und die Maschine neu eingestellt) wird. Das scheint vertretbar. Aber: Wenn die Produktion bereits auf dem 10%-Ausschussniveau läuft, beträgt die Wahrscheinlichkeit nicht einmal ganz 0,80, dass dies zu einem Stichprobenergebnis $X \geq 8$ führt, sodass die Produktion gestoppt würde. Die unterschiedlichen Fehler heißen auch Risiken der Entscheidungsregel, wir sprechen vom Risiko des Typs I und II (siehe Tab. 1). Unter Risiko versteht man aber auch die Wahrscheinlichkeit, einen Fehler des entsprechenden Typs zu begehen. Unsere Entscheidungsregel hat ein Risiko von knapp 5% für einen Typ-I-Fehler und ein Risiko von 20% für einen Typ-II-Fehler. Während das erste Risiko vertretbar erscheint, ist das zweite Risiko völlig inakzeptabel. Abhilfe kann hier eine größere Stichprobe schaffen (Etwa: Inspiziere 400 Stück und unterbreche die Produktion, wenn es 27 und mehr defekte Teile gibt).

Tab. 1: Die Entscheidungsregel als statistischer Test: Fehlerarten und Risiken der Entscheidung.

Status	*Entscheidung*	
	T_0: nicht ablehnen ("annehmen") von H_0	T_1: ablehnen von H_0
H_0: $p_0 = 0.04$ Produktion ist unter Kontrolle	Entscheidung korrekt	Typ-I-Fehler; α-Fehler H_0 wird irrtümlich abgelehnt Risiko 1. Art Produktion wird gestoppt, obwohl alles ok
H_1: $p_1 = 0.10$ Produktion ist außer Kontrolle	Typ-II-Fehler; β-Fehler H_0 wird nicht abgelehnt, obwohl H_1 wahr ist Risiko 2. Art Schwere Störung wird nicht wahrgenommen	Entscheidung korrekt

Im Rahmen des Beispiels wird klar, dass es sich bei den Risiken für Fehlentscheidungen um *bedingte* Wahrscheinlichkeiten und nicht um absolute Wahrscheinlichkeiten handelt. Es kann mit Risiko nicht eine tatsächliche Wahrscheinlichkeit für eine Fehlentscheidung gemeint sein. Es macht keinen Sinn, das Risiko eines Typ-I-Fehlers sehr klein zu machen, damit man sehr selten Fehlentscheidungen trifft. Beide Risiken erhalten ihre Bedeutung erst im Zusammenspiel miteinander. Verringern des Typ-I-Risikos vergrößert nämlich das Typ-II-Risiko, sodass sich die globale Rate von Fehlentscheidungen (das sind *alle* Fehlentscheidungen; also nicht nur jene, die sich auf eine bestimmte Konstellation beziehen) gerade durch die getroffene Maßnahme (verkleinern des Typ-I-Risikos) vergrößern kann.

1.3 Risiko und kleine Wahrscheinlichkeiten

Eine essentielle Deutung für Wahrscheinlichkeit ist die relative Häufigkeit in wiederholten, unabhängigen Versuchen. In diesem Sinn entspricht eine Wahrscheinlichkeit von 0,40 in etwa einer relativen Häufigkeit von 40% in vielen Versuchen. Bei 1000 Versuchen kann man etwa sagen, dass die Ergebnisse etwa einen Spielraum von 3 Prozentpunkten haben; d.h., dass die relative Häufigkeit im

Wesentlichen zwischen 37% und 43% liegen wird; nur sehr selten wird eine stärkere Abweichung vorzufinden sein. Für sehr kleine Wahrscheinlichkeiten ist eine Häufigkeitsdeutung allerdings obsolet. Wir haben dazu einfach keine Daten; die Werte für kleine Wahrscheinlichkeiten entstammen Modellen oder noch besser Szenarien. Im Wesentlichen unterscheiden sich Modelle und Szenarien dadurch, dass ein Modell immer dazu gedacht ist, gut auf die Wirklichkeit zu passen und dass man ein Modell verbessern kann und soll, während ein Szenario einfach ein Durchspielen von Bedingungen ist, auf der Basis „Was wäre, wenn das und das erfüllt ist?" Für eine Erläuterung des Begriffes Szenario verweisen wir auf Borovcnik (2009).

Für BSE, eine Rinderkrankheit mit sehr langer Inkubationszeit, stellen Dubben & Beck-Bornholdt (2010) fest, dass damals alle positiv getesteten Rinder falsch positiv gewesen sein könnten. Wir werden es niemals wissen. Die Sensitivität (0,99) und die Spezifität (0,997) wurden nämlich in einer Studie mit 300 bzw. 1000 Tieren unter *Laborbedingungen* ermittelt (d.h. geschätzt). Mit solchen Diagnoseverfahren sind Risiken für Fehlentscheidungen verbunden. Ja, überhaupt ist die Einführung eines solchen Diagnoseverfahrens mit Risiken verbunden. Wie das BSE-Beispiel zeigt, könnte es sogar sein, dass alle positiv getesteten Rinder falsch-positiv waren. Die Crux an der Sache ist nämlich die überaus kleine Wahrscheinlichkeit für die Prävalenz von BSE. Wie groß diese wirklich ist, darüber gibt es nur Spekulationen. Ein kleines Beispiel (Borovcnik, 2012) soll die Schwierigkeit des Schätzens solch kleiner Wahrscheinlichkeiten illustrieren und aufzeigen, dass die unterstellten Werte keinesfalls aus Daten stammen können.

Aufgabe 2 (Schätzung kleiner Wahrscheinlichkeiten): Eine Wahrscheinlichkeit p für eine bestimmte Eigenschaft E von der Größenordnung 10^{-4} wird aus den Daten aus einer Stichprobe vom Umfang 10000 geschätzt. Wir haben folgende Wahrscheinlichkeiten für die verschiedenen Ergebnisse und den damit verbundenen Schätzwerten:

Tab. 2: Verteilung der Schätzwerte für die unbekannte Wahrscheinlichkeit aus Stichproben vom Umfang 10000.

Wahrscheinlichkeit	*Anzahl der Elemente in der Stichprobe mit E*	*Schätzwert für p*
36,8%	*Keines*	$\hat{p} = 0.0000$
36,8%	*Eines*	$\hat{p} = 0.0001$
18,4%	*Zwei*	$\hat{p} = 0.0002$
6,1%	*Drei*	$\hat{p} = 0.0003$
1,9%	*Vier und mehr*	$\hat{p} \geq 0.0004$

Es gibt demnach ein beträchtliches Risiko (fast 37%), die unbekannte Wahrscheinlichkeit mit null zu schätzen; ein Schätzwert vom Drei- bis Vierfachen des tatsächlichen Werts hat immerhin ein Risiko von 8%. Dazu muss man festhalten, wer verfügt schon über 10000 Daten und kann sicher sein, dass diese Daten aus einem Stichprobenprozess entstammen, d.h., dass sie tatsächlich unabhängig und mit gleichen Wahrscheinlichkeiten aus einer Grundgesamtheit gezogen wurden. Da sind die Erhebungsfehler einfach viel größer als man vermuten würde.

Solche und noch kleinere Wahrscheinlichkeiten (10^{-30} etwa) haben nichts mit der realen Welt und relativen Häufigkeiten zu tun. Sie sind vielmehr aus Szenarien unter sehr einschränkenden Bedingungen gewonnen. Die Werte sind nicht willkürlich gewählt, sondern entsprechen Expertenwissen. Die Frage bleibt, wie kann man Experten validieren. Eine empirische Überprüfung durch relative Häufigkeiten in einer Stichprobe bietet jedenfalls keine Hilfe dabei.

1.4 Drei Wege, Wahrscheinlichkeit begrifflich festzulegen

Es gibt viele Wege, Wahrscheinlichkeiten begrifflich festzulegen; von der großen Vielfalt haben es jedoch nur die folgenden drei geschafft, eine weite Anerkennung und Anwendungsbreite zu entfalten. Wir setzen Akronyme zur Präzisierung dahinter und verfolgen damit die Absicht, die Begriffe gedanklich voneinander abzugrenzen (siehe Borovcnik & Kapadia 2014).

i. Frequentistische Wahrscheinlichkeit (frequentist theory, FQT): Die intuitive Idee dahinter ist, dass relative Häufigkeiten gegen die zugrunde liegende Wahrscheinlichkeit konvergieren, wenn die Versuche unendlich oft wiederholt werden.

ii. Subjektivistische Wahrscheinlichkeit (subjectivist theory; SJT): Abtausch von Geld und Sicherheit
Die Schlüsselidee dahinter ist, dass Wahrscheinlichkeit der Grad des Vertrauens einer Person in eine ungewisse Aussage ist.

iii. Laplacesche Wahrscheinlichkeit (A priori theory; APT): Die Annahme gleicher Wahrscheinlichkeiten für alle Elementarereignisse legt alle anderen Wahrscheinlichkeiten fest. Die Idee dahinter ist, die Gewichte (Wahrscheinlichkeiten) für irgendwelche Information an idealen Glückspielgeräten zu kalibrieren und mitteilbar zu machen.

2. Unbekannte Wahrscheinlichkeiten aus Stichproben schätzen

Wir analysieren die frequentistische Deutung von Wahrscheinlichkeit (FQT) anhand eines Beispiels aus Batanero & Borovcnik (2016). Ziel ist es, das Risiko einschätzen zu lernen, inwieweit eine Stichprobe eine Beobachtung (eine relative Häufigkeit) liefert, welche zu weit weg von der unbekannten Wahrscheinlichkeit ist (also einen bestimmten, vorgegebenen Fehler überschreitet). Ferner soll man an dieser Aufgabe erkennen, dass dieses Risiko mit zunehmendem Umfang der Stichprobe kleiner wird.

2.1 Das Design der Aufgabe

Aufgabe 3 (Schätzen der Wahrscheinlichkeit von Kopf): Jeder Schüler wirft eine Münze 100 Mal (1 steht für Kopf, 0 für Zahl). Wir analysieren das Protokoll (siehe Tab. 3) so:

- Wir teilen die 100 Daten in 20 Blöcke zu 5 (Nr. 1-5, 6-10 usw.) und bestimmen jeweils die Anzahl der Köpfe (0, 1, ..., 5). Dann berechnen wir den Anteil der Köpfe (die relative Häufigkeit) in jedem Block (in jeder Stichprobe der Länge 5) und schätzen damit die unbekannte Wahrscheinlichkeit p für Kopf.
- Wir vereinen zwei aufeinander folgende 5er-Blöcke zu einem 10er-Block und schätzen in jedem 10er-Block (in jeder Stichprobe vom Umfang 10) die unbekannte Wahrscheinlichkeit p für Kopf.
- Wir vereinen zwei aufeinander folgende 10er-Blöcke zu einem 20er-Block; wir zeichnen ein Stabdiagramm der Schätzungen auf der Basis von 20 Würfen.
- Wir vergleichen die Ergebnisse aus 5er-Blöcken innerhalb einzelner Studenten und versuchen, darin ein allgemeines Muster zu erkennen.
- Dann vereinen wir die Daten von (mindestens) 10 Studenten und bestimmen die empirische Verteilung für die Schätzungen nach Blockgröße getrennt. Die Häufigkeitsverteilung der Schätzwerte von 10 Schülern für Stichproben vom Umfang 5 ist in Tab. 3 in der rechten Spalte enthalten. Schließlich zeichnen wir Stabdiagramme mit der Verteilung der Schätzwerte. Wir vergleichen die Stabdiagramme für die unterschiedlichen Stichprobengrößen (siehe Abb. 1): Sieht man ein Muster? Wie steht es mit der Präzision der Schätzung?

2.2 Analyse der Daten

Wir bestimmen jeweils aus der Anzahl der Köpfe in der jeweiligen Stichprobe den Anteil bzw. die relative Häufigkeit der Köpfe, weil dieser Wert die (als unbekannt unterstellte) Wahrscheinlichkeit der Münze für Kopf schätzen lässt. Mit anderen Worten, wir untersuchen die Genauigkeit der Schätzung. Jeder Schüler für sich erkennt, dass seine Schätzwerte variieren; in den ersten vier 5er-Stichproben erhalten wir 3, 1, 3 und 4 Köpfe und damit die Schätzwerte 3/5 = 0,60, 0,20, 0,60 und 0,80 (Tab. 3, links). Aus dem gesamten Protokoll eines Schülers wird ersichtlich, dass die Schätzwerte bei seinen 10 Stichproben vom Umfang 10 etwas weniger variieren und dass die Verringerung in der Variabilität beim Übergang von 10 auf 20 Daten im Block (in der Stichprobe) noch einmal stattfinden sollte.

Tab. 3: Links: Protokoll eines Schülers. Rechts: Häufigkeitsverteilung von 10 Schülern für 5er-Blöcke (5er-Stichproben).

Nr.	Zufall	Ergebnisse im Block		
		5	10	20
1	0			
2	0			
3	1			
4	1			
5	1	3		
6	0			
7	0			
8	1			
9	0			
10	0	1	4	
11	0			
12	1			
13	1			
14	0			
15	1	3		
16	1			
17	1			
18	1			
19	0			
20	1	4	7	11

Blöcke der Länge 5		Häufigkeiten	
Schätzung	Köpfe	absolut	relativ
0.00	0	7	0.035
0.20	1	27	0.135
0.40	2	62	0.310
0.60	3	67	0.335
0.80	4	32	0.160
1.00	5	5	0.025
alle		200	1.000

Vergleichen Schüler ihre Ergebnisse untereinander, so sollten sie erkennen, dass sich ihr Muster nicht immer im Muster der Protokolle der anderen wiederspiegelt. Aber: Wenn alle ihre Daten assemblieren, so ergibt sich ein Muster ähnlich wie in Abb. 1, wenn sie die Häufigkeitsverteilung für jede Stichprobengröße getrennt bestimmen. Zuvor bilden wir die Häufigkeitsverteilung zur jeweiligen Stichprobengröße wie in Tab. 1 rechts für 5er Stichproben. Danach hat sich in 27 der 200 5er-Stichproben von 10 Schülern insgesamt ein Schätzwert von 0,20 (und damit ein großer Schätzfehler ergeben); das sind immerhin in 13,5% aller Stichproben.

Wichtig im Experiment ist auch, dass jeder einen persönlichen Bezug zu seinen Daten hat und „sich" in den Daten der anderen wiederfindet. Das ginge bei einem weniger komplexen Design mit Simulation mit Zufallszahlen am Computer nicht so gut.

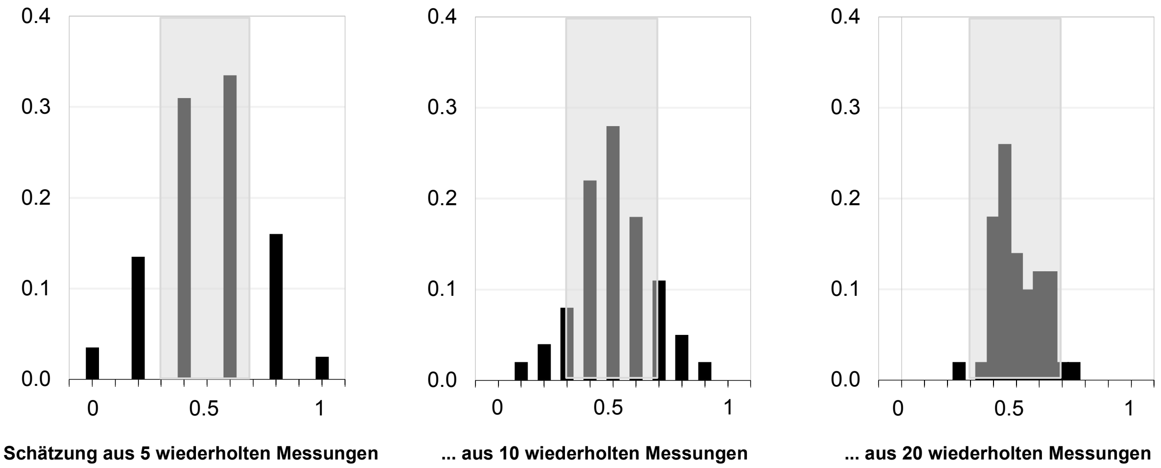


Abb. 1: Verteilung der Schätzwerte mit zunehmendem Stichprobenumfang –
Werte innerhalb des markierten Fensters 0,50±0,20 entsprechen einem Schätzfehler von maximal 0,20.

2.3 Was man aus der Aufgabe lernen soll

Es geht nicht um die Darstellung eines fiktiven Grenzwerts der relativen Häufigkeiten (Gesetz der Großen Zahlen). Dazu reicht die Visualisierung einer endlichen Stichprobe, auch wenn sie noch so groß ist, nicht aus. Vielmehr soll die Qualität eines Sachverhalts nahe gelegt werden: die Verteilung der Schätzwerte wird mit zunehmendem Stichprobenumfang schmäler; d.h., es fallen immer weniger Blöcke mit ihren Schätzwerten außerhalb des zentralen (in Abb. 1 markierten) Fensters. Als Gedankenexperiment ergänzen wir: diese Monotonie kann man extrapolieren; letztlich zieht sich die Verteilung ganz in das Fenster zurück. Wenn die Fensterbreite kleiner gewählt wird, so wird ein analoges Verhalten zu beobachten sein, nur etwas später, d.h. erst bei größeren Stichproben.

3. Wahrscheinlichkeit und Nutzen

Kahneman & Tversky (1979) führten Experimente zum Verständnis von Wahrscheinlichkeit mit Erwachsenen durch. Sie interpretierten deren Verhalten als Fehlvorstellungen und spekulierten über Strategien dahinter: Repräsentativität, Verfügbarkeit, Ankern etc. Sie zeigten, wie inkonsistent Leute bei Ungewissheit sind, speziell wenn die Folgen von Entscheidungen die Situation mitbestimmen.

3.1 Verhalten in Gewinn- und Verlustsituationen

Folgende Beispiele zeigen, wie schwer sich Menschen tun, ihre Wahrnehmung und Einschätzung von Nutzen bzw. Schaden, der mit den Möglichkeiten verbunden ist, von der Einschätzung der Wahrscheinlichkeit der Möglichkeiten zu trennen (Tversky & Kahneman, 1981).

Tab. 4: Gewinn- und Verlust der Optionen in den beiden Experimenten (Kahneman & Tversky, 1979) und ihren Varianten.

	K/T-Sicht		Andere Sicht (wir)		Viele Einzelentscheidungen	
	Experiment 1:		Experiment 1^*:		Experiment 1^s:	
Zukunft [$]	Option a_1	Option a_2	a_1^* Tue nichts	Option a_2^*	Option b_1	Option b_2
	1000	2500 mit ½ 0 mit ½	1000 die hat man ja	1500 mit ½ –1000 mit ½	1	2,5 mit ½ 0 mit ½
	Experiment 2:		Experiment 2*:		Experiment 2^s:	
Zukunft [$]	Option a_3	Option a_4	a_3^* Tue nichts	Option a_4^*	Option b_3	Option b_4
	–1000	–2500 mit ½ 0 mit ½	–1000 die hat man ja	–1500 mit ½ 1000 mit ½	–1	–2,5 mit ½ 0 mit ½

In Gewinnsituationen sind Menschen risikoscheu: In Gewinnsituationen wie in Experiment 1 ziehen viele Menschen die Option mit den sicheren $ 1000 vor, auch wenn a_2 einen *Wert* von 1250 hat.

In Verlustsituationen suchen Menschen das Risiko: In Verlustsituationen wie in Experiment 2 ziehen viele Menschen die Option a_4 vor; d.h., sie suchen das Risiko, statt auf dem sicheren Verlust von –1000 stehen zu bleiben.

Von Kahneman & Tversky wird das Verhalten in Spalte 1 von Tab. 4 als risiko-aversiv (Experiment 1, Gewinnsituationen) und risiko-suchend (Experiment 2, Verlustsituationen) charakterisiert – jedenfalls aber als Fehleinschätzung gemäß dem erwarteten Gewinn/Verlust. Es lohnt sich aber, die Situation umzudeuten (Spalte 2). Denn eigentlich ist die jeweils sichere Optionen (a_1 bzw. a_3) vergleichbar mit dem Status quo, das, was man ohnehin schon erreicht hat. D.h., darüber gibt es weder Unsicherheit noch eine Option. Es beschreibt lediglich den derzeitigen Zustand. Dann wird aber die Möglichkeit, doch etwas zu tun, immer mit diesem Ausgangszustand verglichen, als ob dieser real wäre. Hierbei sieht die Situation ganz anders aus und das Verhalten, das Kahneman & Tversky beobachtet haben, kann nicht mehr als Abweichung von Rationalität gedeutet werden (als risikoscheu bzw. als risikosuchend) sondern folgt anderen, verständlichen psychologischen Mustern.

Die Person hat die $ 1000 schon. Der zusätzliche Betrag ist nur $ 1500; dabei besteht aber ein Risiko, die 1000 wieder zu verlieren und auf 0 zu gehen! Es lohnt sich nicht, die 1000 aufs Spiel zu setzen für den kleinen zusätzlichen Gewinn von 1500. Die Situation ist eher als Verlustsituation aufzufassen. Eine Person, die bereits 1000 hat, muss besser bezahlt werden, um das Risiko zu suchen, alles wieder zu verlieren. Menschen, die bereits Schulden haben wie in Experiment 2^*, werden größere Risiken eingehen, um diese loszuwerden und a_4^* wählen. Die Situation in Experiment 2 kann vielmehr als Gewinnsituation beschrieben werden. Wird eine Option ohne Zufall dargestellt, tendieren Menschen dazu, einen relativen Vergleich anzustellen. Der führt i.A. zu anderen Entscheidungen.

3.2 Logik wiederholter einmaliger Entscheidungen

Es klingt paradox, dass eine Entscheidung, welche wiederholt wird, anderen Gesetzmäßigkeiten folgen soll als eine Einzelentscheidung. Was optimal ist für *eine* Entscheidung, muss nicht mehr gut sein, wenn diese Entscheidung oft wiederholt wird. Wenn wir *einmal* entscheiden, so entscheiden wir in der anderen Sicht (*wir*) anders als Kahneman & Tversky (*K/T*); die präferierte Entscheidung sei so notiert *K/T* (*wir*); in der Gewinnsituation also b_2 (b_1); in der Verlustsituation dagegen b_3 (b_4).

Wenn wir 1000 Mal spielen, gewinnt man mit Option b_1 $ 1000. Um in Experiment 1^s mit b_2 mehr als $ 1000 zu gewinnen, muss man mehr als 400 Mal gewinnen; die Wahrscheinlichkeit dafür ist $1–10^{-10}$; wir gewinnen sogar mindestens $ 1250 (dazu muss man 500 Mal oder öfter gewinnen) mit mehr als 0,51 Wahrscheinlichkeit. In der Verlustsituation von 2^s gilt: Mit b_4 verliert man ziemlich sicher mehr als $ 1000 (das ist der Verlust mit b_3). Wenn eine Person viele gleichartige Entscheidungen trifft, so kann sie nach ganz anderen Gesichtspunkten entscheiden. Etwa hat eine Versicherung viele Klienten und damit viele Einzelentscheidungen, während der einzelne Versicherungsnehmer nur einen Vertrag abschließt. Die Logik der wiederholten Einzelentscheidungen orientiert sich an anderen Gesichtspunkten als eine einzige Entscheidung.

3.3 Widersprüche, Auffassungsunterschiede, Nutzen & Wahrscheinlichkeit

Die Situation wird schwieriger und das Verhalten mag beeinflusst werden, wenn der Nutzen des Geldes ins Spiel kommt; wenn die Beträge viel größer werden; wenn man die Ausgangssituation anders auffasst (wie soeben).

Man kann das Verhalten der Leute, die a_2^* vermeiden und a_4^* suchen, so interpretieren, als ob sie es vermeiden, zusätzlich 250 zu gewinnen, aber unbedingt weitere 250 verlieren „wollen“ (wie dies Kahneman & Tversky tun). In der Versuchssituation sehen sich die Leute aber mit einer anderen Ausgangssituation konfrontiert (1.000 haben oder Schulden von 1.000 haben). Ihr Verhalten ist nicht inkonsistent sondern dieser Auffassung der Situation geschuldet. Die Entscheidung ist wieder anders, wenn es sich um eine wiederholte Entscheidung handelt, wie wir gesehen haben. Die Entscheidung wird auch anders, wenn die Gewinne bzw. Verluste viel höher werden, aber die Wahrscheinlichkeit jeweils sehr klein ist (es war schon die Rede von der Problematik kleiner Wahrscheinlichkeiten, die nur schwer durch relative Häufigkeiten validiert werden können): i. Wenn “Tue nichts” mit einem potentiell hohen (aber unwahrscheinlichen) Gewinn verglichen wird, welcher nur wenig kostet – Lotto. ii. Wenn mit einem potentiell hohen (aber unwahrscheinlichen) Verlust verglichen wird, was aber relativ viel kostet – Versicherungsvertrag. Gigerenzer (2007) untersucht den Effekt von Bauchentscheidungen, die der Komplexität der Situation geschuldet sind; gemessen an der langen Dauer der Informationsbeschaffung sind solche Entscheidungen eigentlich gar nicht so schlecht.

3.4 Odds – eine bequeme Art, über Wahrscheinlichkeiten zu sprechen

Bevor Wahrscheinlichkeiten als ein Maß für das Eintretens eines Ereignisses als eine Zahl zwischen 0 und 1 eingeführt wurde (bei Jakob Bernoulli), hat man die Möglichkeit des Eintretens immer im Vergleich zueinander angegeben; speziell hat man davon gesprochen, dass die Chancen des Eintretens im Vergleich zum Ausbleiben besser als 50:50 stehen; man hat als implizit ein Chancenverhältnis angegeben. Das ist beim Wetten auf das Ereignis (die Aussage) E auch heute noch üblich. Man sagt, die Chancen (Odds) stehen $p : q$.

Odds sind ein Verhältnis der Wahrscheinlichkeiten, dass E eintritt im Vergleich dazu, dass E ausbleibt, i.e., $P(E):P(\overline{E})$. Die Odds für einen Sechser beim Würfeln stehen 1:5. Aus Odds kann man die Wahrscheinlichkeit zurückrechnen; es gilt: $P(E)=\frac{p}{p+q}$. Konkret: $P(Sechser)=\frac{1}{1+5}=\frac{1}{6}$. Wir verwenden Odds auch zum Kalibrieren (Tunen) unserer persönlichen Wahrscheinlichkeit.

4. Paradigmatische Beispiele

Wahrscheinlichkeit ist aus dem Glücksspiel und den Versicherungen gewachsen und somit von den Wurzeln her mit dem Begriff Risiko verbunden. Im Folgenden werden anhand einiger exemplarischer Beispiele die Konstituenten der Situation begreifbar: der Versicherungsvertrag; Szenarien, die zeigen, dass unterschiedliche Interessensgruppen unterschiedliche Sichtweisen haben müssen; Entscheidungen hängen von Kriterien ab (es gibt keine allgemein gültige Optimierung); bedingte Wahrscheinlichkeiten reflektieren einen bestimmten Informationsstand.

4.1 Abtausch von Geld gegen Sicherheit

Aufgabe 4 (Versicherung): Ist es besser, eine Vollkasko-Versicherung für das neue Auto für das kommende Jahr abzuschließen oder nicht? Im Nachhinein ist die Frage leicht zu beurteilen: ist kein Unfall passiert, ist keine Versicherung (a_2) die bessere Entscheidung. Wir müssen aber im Vorhinein im Angesicht der Unsicherheit entscheiden. Orientiert man sich am maximalen Verlust (siehe Tabelle 5), dann ist Versicherung abschließen (a_1) besser. Das spiegelt risikoscheues Verhalten.

Tab. 5: Gewinn- und Verlust der Optionen Versicherung ja oder nein in einem vereinfachten Szenario für die Zukunft.

	Kosten (€)	a_1 = Versicherung ja	a_2 = Keine Versicherung
Potentielle Zukunft	t_1 = Kein Unfall	1.000	0
	t_2 = Totalschaden	1.000	30.000

Die Versicherungsgesellschaft nützt Information aus Unfallstatistiken.

- Sie basiert ihr Modell auf Geld und eine Schätzung der Wahrscheinlichkeit für Schäden (und damit verbundene Zahlungen) durch die relative Häufigkeit der Ereignisse in der Vergangenheit.
- 2% Totalschaden: $30.000\times0{,}02 + 0\times0{,}98 = 600$ plus Spesen und Profit.

Der Autobesitzer kann die Unfallstatistiken nicht nutzen.

- Er muss sich auf persönliche Risiken beziehen: Fahrkünste, riskantes Verhalten, wie viel er fährt, wo und wann (Autobahn, Winter, Nacht etc.)?
- und muss den Nutzen des Geldes berücksichtigen.

Stehen die Odds für Totalschaden auf 1:9, dann ist die implizite Wahrscheinlichkeit dafür gleich 1/10. Das ergibt erwartete Kosten für a_2 von $30.000\times0{,}1+0\times0{,}9 = 3.000$; für die Option a_1 betragen die Kosten immer 1.000. Daher ist die Versicherungspolizze besser. Wenn man aber eine kleinere Wahrscheinlichkeit für den Totalschaden unterstellt, dann wird die Entscheidung kippen auf keine Versicherung. Knackpunkt: wie "misst" man seine persönlichen Wahrscheinlichkeiten?

Die Versicherung ist bereit, das Risiko der finanziellen Folgen eines solchen Unfalls auf sich zu nehmen, wenn sie dafür einen bestimmten Betrag im Vorhinein bekommt. Im Kontrakt zahlt der eine, um Unsicherheit (möglicher Verlust) zu vermeiden, der andere wird dafür bezahlt, dass er die Position der Sicherheit (kein Verlust) aufgibt. Paradox: Wie kann eine Entscheidung gut für beide sein? Das kann man verstehen, wenn die Versicherung in Geld, der Kunde aber in Nutzen denkt. Es kann aber auch sein, dass beide ein anderes Kriterium der Entscheidung zugrunde legen: die Versicherung kann aus der Unfallstatistik eine Prognose für den Gewinn erstellen; der Kunde orientiert sich eventuell am schlimmsten Fall (der Totalschaden), den er unbedingt vermeiden will (muss). Darüber hinaus haben beide eine unterschiedliche Wahrscheinlichkeit für den Totalschaden: die Versicherung hat eine frequentistische Deutung aus der Unfallstatistik (FQT), der Kunde kann nur seine persönlichen Chancen zu bewerten versuchen (SJT).

Folgende Überlegung kann den Prozess der Bewertung der persönlichen Chancen unterstützen: Für welche Odds kippt die Entscheidung von *Versicherung* zu *keine Versicherung*? Sind die Odds für Unfall kleiner als 1:29 (also etwa 1:50), dann ist "keine Versicherung" besser; sind sie größer (also etwa 1:10), so ist „Versicherung" die bessere Option. In diesem Sinne muss man die genauen Chancen gar nicht kennen, sondern nur mit dieser Marke vergleichen. Bewertet man die finanziellen Folgen zusätzlich mit Nutzen, so kann man nachvollziehen, warum sich so viele Leute versichern lassen (30.000 im Schadensfall zu verlieren ist – in Nutzen gemessen – viel mehr als das 30-fache des Verlustes von 1.000, was die Versicherung ausmacht).

4.2 Wahrscheinlichkeiten nützen, um Entscheidungen zu treffen

Aufgabe 5 (Optimierung der Auflage): Welche Auflage optimiert den Gewinn? Die Nachfrage N für eine Wochenzeitschrift hänge von zufälligen Faktoren ab; die Verteilung der Nachfrage sei durch Marktbeobachtung bekannt. Der Preis eines Heftes beträgt € 1,60. Wie viele Hefte soll man drucken, wenn man die Wahl zwischen 1000, 2000, …, 5000 hat?

Tab. 6: Verteilung der Nachfrage sowie Kosten der verschiedenen Optionen bezüglich der Auflage.

Nachfrage n_i	1000	2000	3000	4000	5000
Wahrscheinlichkeit p_i	0,40	0,30	0,20	0,06	0,04
Auflage a_j	1000	2000	3000	4000	5000
Kosten $C(a_j)$	2000	2200	2400	2600	2800

Wir führen die Berechnungen nur in Geld und nicht in Nutzeneinheiten durch. Zunächst berechnen wir die Verteilung des Netto-Gewinns bei einer Entscheidung für 2000 Kopien (siehe Tab. 7).

Tab. 7: Verteilung des Nettoprofits bei Auflage 2000.

Entscheidung		a_2	2.000
		Kosten	2.200
Nachfrage	Wahrsch.	Umsatz	Nettoprofit
1.000	0,40	1.600	-600
2.000	0,30	3.200	1.000
3.000	0,20	3.200	1.000
4.000	0,06	3.200	1.000
5.000	0,04	3.200	1.000

Wir sehen, dass ein Verlust von –600 eintreten kann; der Erwartungswert des Gewinns beträgt 360. Wir notieren die Verteilung des Gewinns für die anderen Entscheidungen kompakt in den Spalten einer so genannten Gewinn-Matrix (Tab. 8). Aus der ersten Spalte sehen wir, dass eine Entscheidung für Auflage 1000 immer, egal, wie hoch die Nachfrage ist, einen Verlust von –400 bringt. Offensichtlich wird man eine solche Entscheidung vermeiden wollen. Allerdings ist der mögliche Verlust bei

Auflage 2000 noch höher (–600), bei 3000 Stücken steigt das Risiko (im Sinne der Höhe des Verlusts) auf –800 an; es lockt allerdings ein erwarteter Gewinn von 640, der deutlich besser ist als bei Auflage 2000. Auflage 4000 birgt ein noch höheres Risiko und der erwartete Gewinn geht zurück auf 600, was immer noch sehr viel besser als bei Auflage 2000 ist, aber mit einem möglichen Verlust von –1000 verbunden ist. Orientiert man sich am maximalen Verlust, dann ist die Auflage 1000 die beste. Die nach diesem Kriterium beste Entscheidung ist aber in diesem Fall völlig inakzeptabel. Orientiert man sich am erwarteten Gewinn, so ist die Entscheidung für eine Auflage 3000 am besten.

Tab. 8: Verteilung des Nettoprofits bei allen Auflagen.

Matrix der Profite *U*		**Entscheidung: Anzahl der Kopien a_j**					Preis / Heft
	a_j	1.000	2.000	3.000	4.000	5.000	1,60
	Kosten C (a_j)	2.000	2.200	2.400	2.600	2.800	
	n_i						p_i
Nachfrage n_i	1.000	-400	-600	-800	-1.000	-1.200	0,40
	2.000	-400	1.000	800	600	400	0,30
	3.000	-400	1.000	2.400	2.200	2.000	0,20
	4.000	-400	1.000	2.400	3.800	3.600	0,06
	5.000	-400	1.000	2.400	3.800	5.200	0,04
erwarteter Profit (a_j)		-400	360	640	600	464	
maximaler Verlust (a_j)		-400	-600	-800	-1.000	-1.200	

Mit dem Minimax-Prinzip kann man eine Entscheidung finden, ohne Wahrscheinlichkeiten zu verwenden; das Prinzip kann aber zu unsinnigen Entscheidungen führen. Modelliert man die Nachfrage mit Wahrscheinlichkeiten (etwa von relativen Häufigkeiten aus der Vergangenheit), so kann das Kriterium Erwarteter Profit angewendet werden. Das führt i.A. zu besseren Lösungen, birgt aber das Risiko höherer Verluste in sich. Man kann wieder mit Nutzen statt mit Geld rechnen (der Nutzen ist im Allgemeinen eine nicht-lineare Funktion des Geldes und kann sich von Person zu Person stark unterscheiden). Verschiedene Stakeholder präferieren jeweils andere Kriterien (Borovcnik & Kapadia, 2011b).

4.3 Bedingte Wahrscheinlichkeiten: Zufällig aus einer dynamischen Urne ziehen

An der folgenden Aufgabe zur Diagnose von Krankheiten wird zunächst der Begriff der bedingten Wahrscheinlichkeit klarer als sonst darstellbar; im nächsten Abschnitt wird der Kontext der Aufgabe genützt, um im Kontext der Diagnose die damit verbundenen Risiken durchschaubarer zu machen.

Aufgabe 6 (Medizinische Diagnose): Ein medizinischer Test (Labortest, Röntgen etc.) heiße positiv, wenn er auf eine Krankheit *K* hinweist, im anderen Fall heiße das Ergebnis negativ. Wir unterstellen eine Studie, in der der Status der Krankheit zweifelsfrei festgestellt werden konnte und das Testergebnis bekannt ist. Wir interpretieren nun diese Häufigkeitsinformation (FQT, welche die unbekannten Wahrscheinlichkeiten schätzt) als gleiche Wahrscheinlichkeiten (APT), indem wir für jede Person eine Kugel in eine Urne legen. Wir haben 108 rote und 892 blaue Kugeln (für den Status der Diagnose). Die Kugeln seien öffenbar; innen sei ein Zettel mit dem Status der Krankheit. Es enthalten 9 der roten Kugeln einen Zettel *K*, 891 der blauen Kugeln einen Zettel *nicht-K*.

Tab. 9: Fiktive Studie über gemeinsames Auftreten einer Krankheit mit den medizinischen Befunden positiv und negativ.

Testurteil	Status der Krankheit *K*	*Nicht-K*	Alle
Positiv +	9	99	108
Negativ –	1	891	892
Alle	10	990	1.000

„Daten“ aus einer Studie der Krankheit *K*.

Zufällig Ziehen aus einer Urne mit dieser Besetzung

Wir können fragen: Wie groß ist die Wahrscheinlichkeit, dass a) eine rote Kugel gezogen wird? b) …, dass die Kugel rot ist, wenn ein K-Zettel drinnen ist? c) …, einen Zettel K zu finden, wenn die gezogene Kugel rot ist? In unserem Setting erhalten wir folgende Wahrscheinlichkeiten (wir notieren $P_K(+)$ statt $P(+\,|\,K)$, um anzudeuten, dass es sich bei der bedingten Wahrscheinlichkeit nicht um die Funktion P handelt):

a) $P(+) = \frac{108}{1000} = 0{,}0180$, $P(K \cap +) = \frac{9}{1000} = 0{,}0090$.

b) $P_K(+) = \frac{9}{10} = 0{,}9000$ verglichen mit $P_{Nicht-K}(+) = \frac{99}{990} = 0{,}1000$.

c) $P_+(K) = \frac{9}{108} = 0{,}0833$ verglichen mit $P_-(K) = \frac{1}{892} = 0{,}0011$.

Wir stellen fest, dass folgende Beziehung allgemein gilt: $P_+(K) = \frac{9}{108} = \frac{9/1000}{108/1000} = \frac{P(K \cap +)}{P(+)}$.

Das ist genau die übliche Art, die bedingte Wahrscheinlichkeit zu definieren. Wir hatten bislang bedingte Wahrscheinlichkeiten direkt aus dem Kontext bestimmt, weil ohnehin klar war, welche Zahlen heranzuziehen waren. Aufgabe 6 dient zunächst zur Motivation der Definition der bedingten Wahrscheinlichkeit.

5. Methoden zum leichteren Umgang mit Wahrscheinlichkeiten

Wir greifen den Kontext der medizinischen Diagnose in Aufgabe 6 auf, um die verschiedenen Typen von Information (a priori, Daten und a posteriori-Wissensstand) herauszuarbeiten. Diese Typen von Information zu kennen und mit dem Kalkül der Bayes-Formel (der bedingte Wahrscheinlichkeiten in beiden Richtungen miteinander in Verbindung stellt) leichtfüßig umgehen zu können, ist für die Einschätzung der Risiken, die mit solchen Diagnoseverfahren einhergehen, unabdingbar.

5.1 Die relativen Häufigkeiten so lesen, als ob sie Wahrscheinlichkeiten wären

Wie schon in Aufgabe 6 lesen wir die relativen Häufigkeiten, die mit Tabelle 9 verknüpft sind, als ob sie Wahrscheinlichkeiten wären (ohne auf den Schätzfehler einzugehen). Zufällig Auswählen aus einer Urne mit entsprechenden Kugeln rechtfertigt diese Vorgangsweise. Wir interpretieren die verschiedenen Informationen über den Status der Krankheit gleich über die entsprechenden Chancenverhältnisse.

A priori-Wissen spricht gegen *K*, weil die Odds für *K* 1 : 99 betragen:	Der medizinische Test + spricht für *K* mit einem Likelihood-Verhältnis von 9 : 1:	A posteriori-Wissen bei Test + spricht mit Odds von 9 : 99 für *K*:
$P(K) = \frac{10}{1000} = 0{,}01$,	$P_K(+) = \frac{9}{10} = 0{,}9000$	$P(K \cap +) = \frac{9}{108}$
$P(Nicht\text{-}K) = \frac{990}{1000} = 0{,}99$.	$P_{Nicht-K}(+) = \frac{99}{990} = 0{,}1000$	$P(Nicht\ K \cap +) = \frac{99}{108}$
Es gibt 99 Mal mehr *Nicht-K's* als es *K's* gibt.	Der Befund + ist unter *K's* 9 Mal so häufig wie unter *Nicht-K's*	Auf einen mit *K* kommen 11 *ohne K* in der Gruppe aller +

Vorweg, ohne medizinischen Befund, sprechen die Daten stark *gegen* das Vorliegen der Krankheit (Faktor 99); der Befund spricht zwar *für* die Krankheit (mit einem Faktor 9), aber diese Information ist weniger deutlich als das Wissen vorweg. Insgesamt – bei Berücksichtigung beider Informationen – spricht es weniger stark gegen das Vorliegen, aber noch nicht für das Vorliegen der Krankheit. Auch wenn sich das Risiko, die Krankheit zu haben, vergrößert hat, bedarf es noch weiterer Befunde bis zur Klärung der Lage.

5.2 Bayes-Formel aus Kontingenztabellen

Wir tragen in Tabelle 9 zusätzlich zu den Anzahlen noch die entsprechenden Ereignisse (Aussagen) ein, um damit die entsprechenden Formeln bis hin zur Bayes-Formel zu motivieren. Man kann nun aus Tabelle 10 alle interessierenden Wahrscheinlichkeiten direkt abzulesen. Die totale Wahrscheinlichkeit von *Pos* ergibt sich aus der ersten Zeile: $P(Pos) = P(\,Pos \cap K) + P(\,Pos \cap \textit{Nicht-K}\,)$. Nach der Multiplikationsregel ersetzt man diesen Ausdruck durch: $P(Pos \mid K) \cdot P(K) + P(Pos \mid \textit{Nicht-K}) \cdot P(\textit{Nicht-K})$. Die Bayes-Formel liest man aus dieser Zeile als bedingte Zeilenwahrscheinlichkeit von *Krank* unter *Pos* ab, ohne den Nenner zu spezifizieren: $P(K \mid Pos) \;=\; \dfrac{P(Pos \mid K) \cdot P(K)}{\ldots}$. Leute versuchen, diese posteriori-Wahrscheinlichkeit intuitiv zu schätzen. Das geht meist schief.

Tab. 10: Einordnen der Einträge in verschiedene Gesamtheiten – Ablesen wichtiger mathematischer Beziehungen.

Biometrischer Test	*Status der Erkrankung*					
	Krank K		*Nicht-K (Gesund)*		*Alle*	
Positiv +	$Pos \cap K$	9	$Pos \cap \textit{Nicht-K}$	99	*Pos*	108
Negativ –	$Neg \cap K$	1	$Neg \cap \textit{Nicht-K}$	891	*Neg*	892
Alle	*K*	10	*Nicht-K*	990	*Alle*	1000

5.3 Odds, Bayes-Formel und Abwägen der Information (Evidenz)

Die Darstellung der Bayes-Formel kann durch ein Baumdiagramm unterstützt werden. Wir wollen die Bezeichnungen vom Kontext lösen und sie näher zur Situation der statistischen Beurteilung bringen. Vorweg, a priori hat man zwei Hypothesen gegeneinander abzuwägen: $H_1 : K$ und $H_2 : \textit{Nicht-K}$. Der empirische Befund *E* (hier das Ergebnis *Pos* der Diagnose) wird mittels der bedingten Wahrscheinlichkeit $P_E(H_1) = P(H_1 \mid E)$ miteinbezogen; die neue (a posteriori) Information lautet:

$$P_E(H_1) \;=\; \frac{P(H_1) \cdot P_{H_1}(E)}{P(H_1) \cdot P_{H_1}(E) + P(H_2) \cdot P_{H_2}(E)}.$$

Wenn wir Odds betrachten, müssen wir nur das Verhältnis $P_E(H_1) : P_E(H_2)$ berechnen; der komplizierte Nenner kürzt sich heraus:

$$\frac{P_E(H_1)}{P_E(H_2)} \;=\; \frac{P(H_1)}{P(H_2)} \;\times\; \frac{P_{H_1}(E)}{P_{H_2}(E)}$$

$$\underset{\text{relativ auf Daten}}{\textit{Neue} - \textit{posteriori} - \textit{Odds}} \;=\; \underset{\text{ohne Daten}}{\textit{Alte} - \textit{priori} - \textit{Odds}} \;\times\; \underset{\text{Trennkraft der Daten}}{\textit{Likelihoodverhältnis}}$$

Die a posteriori-Odds ergeben sich als Produkt der a priori-Odds mit einem Quotienten, der als Trennkraft der Daten (den Gewichten des empirischen Befundes) interpretiert werden kann. Drei Typen von Information sind allgemein in Bayes-Problemen beteiligt, siehe Tab. 11. Wir können der Reihe nach a priori-Wissen und dann das Wissen aus dem medizinischen Test anwenden und erhalten das a posteriori-Wissen.

Tab. 11: Bewertung der verschiedenen Typen von Informationen in Chancenverhältnissen.

A posteriori-Wissen spricht für *K*	*A priori-Wissen* spricht für (gegen) *K*	*Test* + spricht für *K*
mit Odds von	mit Odds für *K* von	mit einem Likelihood-Verhältnis
9 : 99 = 1 : 11	1 : 99	9 : 1

Es ist hilfreich, die entsprechenden Chancenverhältnisse als Verkleinerungsfaktoren (verkleinert wird in diesem Maßstab die „Sicherheit“ von 1) wie beim Kopieren zu denken. Erst wird mit dem Faktor 1:99 verkleinert, dann mit einem Faktor von 9:1 vergrößert. Es ergibt sich ein Skalierungsfaktor von $\frac{1}{99} \times \frac{9}{1} = \frac{9}{99} = \frac{1}{11}$. Diesem Chancenverhältnis entspricht eine Wahrscheinlichkeit von 1/12.

Es wird unmittelbar klar, warum die Wahrscheinlichkeit für die Krankheit nach positivem Befund so klein ist. Nicht dass der Befund so schlecht ist, er hat ja immerhin das Chancenverhältnis mit dem Faktor 9 vervielfältigt. Nein, es ist das ungünstige Chancenverhältnis für *K* a priori, weil die Krankheit selten ist. Je seltener eine Krankheit ist, desto schwieriger ist es, sie korrekt zu diagnostizieren. Nicht nur, weil es an medizinischer Erfahrung fehlt, sondern weil es bei der Diagnose nur unvollständige Genauigkeit gibt (Sensitivität und Spezifität sind niemals 1). Die Darstellung mit Odds liefert eine einfache Beziehung. Die Skalierungsfaktoren wirken linear, d.h., wird einer etwa verdoppelt, so verdoppeln sich die Chancen a posteriori. Schöner könnte man das Zusammenspiel der unterschiedlichenTypen von Information nicht in Szene setzen. Alle anderen Visualisierungen gehen daran vorbei und treffen den genuin probabilistischen Charakter der Situation nicht.

5.4 Untergruppenanalyse – Schätzen von (bedingten) Wahrscheinlichkeiten

Mit einer Kontingenztabelle sind weitere Häufigkeitsverteilungen verbunden. Der Vergleich der Untergruppen in Tabelle 12 zeigt, welchen Einfluss der Wert des anderen Merkmals hat, der diese Gruppe erzeugt. Etwa ist die Verteilung des Testergebnisses in den Untergruppen, die durch das Vorliegen von *K* erzeugt werden, gänzlich verschieden; selten tritt ein positives Ergebnis bei Gesunden auf, dagegen ist ein positives Ergebnis sehr häufig bei Kranken vertreten. Ein solches Merkmal kann demnach zwischen Gesunden und Kranken trennen, dies umso besser, je weiter die Verteilungen auseinandergehen (wir haben das als Likelihood-Verhältnis bei positivem Test angesprochen).

Tab. 12: Verschiedene Zerlegungen der Gesamtheit in Untergruppen: Zeilen- und Spaltenverteilung.

Mediz. Test	Status		
	K	*Nicht-K*	Alle
Pos +	9	99	108
Neg –	1	891	892
Alle	10	990	1.000

	Werte		
	K	*Nicht-K*	Alle
Gruppe 1 +	8,3	91,7	100,0
Gruppe 2 –	0,1	99,9	100,0
Alle	1,0	99,0	100,0

Spaltenprozent

Zeilenprozent

Medizinischer Test in Untergruppen in %

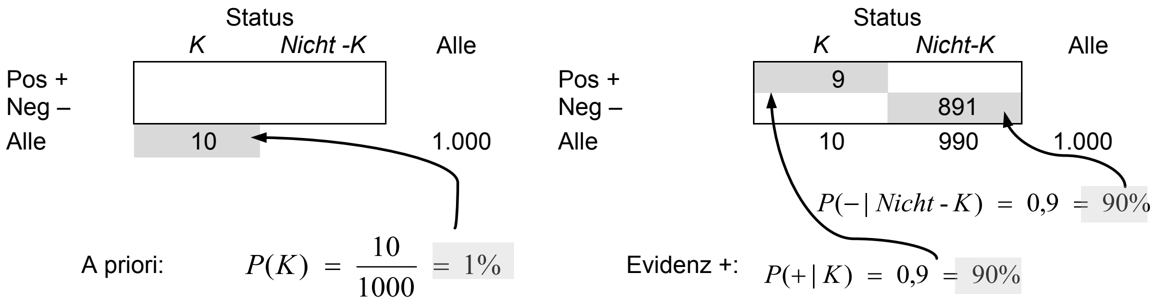
Werte	# 1	# 2	
	K	*Nicht-K*	Alle
Pos +	90,0	10,0	10,8
Neg –	10,0	90,0	89,2
Alle	100,0	100,0	100,0

Mit Häufigkeiten als Schätzungen für die Wahrscheinlichkeiten können wir die Tabellen mit Zeilen- und Spaltenprozenten als bedingte Wahrscheinlichkeiten interpretieren.

5.5 Die Informationen hierarchisch lesen – natürliche Häufigkeiten

Es gibt eine einfache Möglichkeit, die verschiedenen Informationen über (bedingte) Wahrscheinlichkeiten in ganze Zahlen einer Kontingenztabelle umzusetzen. Diese wird sehr übersichtlich, wenn es sich jeweils um binäre Merkmale handelt, wie das im medizinischen Diagnoseproblem der Fall ist. Man muss nur beachten, dass man die Gesamtheit nicht zu groß wählt, weil sonst die Anzahlen unübersichtlich werden. Allerdings treten bei zu kleiner Anzahl Rundungsfehler auf. Wir tun so, als ob sich die Wahrscheinlichkeiten auf ein statistisches Dorf (die Referenzpopulation) beziehen und berechnen erwartete Häufigkeiten nach der Formel $E = Np$. Tab. 13 zeigt, wie man der Reihe nach diese Werte aus gegebener Information der Prävalenz $P(K) = 1\%$, der Sensitivität $P(+ \mid K) = 90\%$ und der Spezifität $P(- \mid Nicht\text{-}K) = 90\%$ berechnen und in die Tabelle eintragen kann.

Tab. 13: Umsetzen von (bedingten) Wahrscheinlichkeiten in erwartete Anzahlen einer Kontingenztabelle: ein statistisches Dorf.



Die Idee ist alt und wurde von Gigerenzer (2002) aufgegriffen; statt von Erwartungswerten ist dort von natürlichen Häufigkeiten die Rede. Nachteil der Veranschaulichung ist, dass Anzahlen eher als Fakten denn als modellabhängige Größen gesehen werden. Während 1% für die Prävalenz durchaus mit Unsicherheit behaftet ist, sind es dann 10 (und nicht 15 oder gar 20). Spiegelhalter & Gage (2015) gehen sogar noch weiter und zeichnen Ikons für die Anzahlen. Ein weiterer Nachteil ist, dass man ob der materiellen Darstellung (die eine Verwechslung mit der Realität provoziert) aus den Augen verliert, dass die bedingten Wahrscheinlichkeiten von der Subgruppe, der man angehört, abhängen. So etwa sind weder Prävalenzen noch Sensitivitäten und Spezifitäten Konstanten sondern nur Schätzwerte, die überdies wesentlich durch diese Subgruppe bestimmt werden. Dann sieht die Kontingenztabelle gänzlich anders aus. Überdies wird nur das Resultat selbst leicht ablesbar, die Gründe, warum das so ist, gehen aus dieser Darstellung nicht hervor. Da sind die an sich komplizierteren Odds viel besser geeignet, den Einfluss verschiedener Informationen auf die endgültige Einschätzung offen zu legen.

6. Konsequenzen für den Umgang mit Risiko

Der Sprachgebrauch von Risiko ist sehr heterogen; der Platz hier reicht nicht aus, deswegen der Verweis auf Borovcnik (2015). Mal sagen wir zur schlechteren der Möglichkeiten (den Schaden) Risiko, mal sprechen wir von der Wahrscheinlichkeit, diesen Schaden zu erleiden. Ein anderes Mal meinen wir beides. Allerdings zeigen die Beispiele, dass in der Wahrnehmung ohnehin eine Trennung schwer möglich ist und es daher zu vielen Fehleinschätzungen kommt (etwa weil man den maximalen Schaden vermeiden möchte). Weiters sind die kleinen Wahrscheinlichkeiten oft sehr abträglich, sich mit Risiken rational auseinanderzusetzen. Dazu fehlen einfach Daten und wir werden diese nie zur Verfügung haben; eine Häufigkeitsdeutung von so kleinen Wahrscheinlichkeiten ist eigentlich nur eine façon de parler. Die folgende Liste fokussiert auf die Brennpunkte der Darstellung:

- Wahrscheinlichkeit ist ein virtuelles Konzept, das sich am besten für Vergleiche eignet: Es braucht verschiedene Optionen, um Wahrscheinlichkeit zu nutzen. Man kann Risiken nur vergleichen.
- Die Entscheidungen sind abhängig von den verwendeten Zielkriterien und der Interpretation von Wahrscheinlichkeit.
- Information ist für die verschiedenen Interessensgruppen unterschiedlich nützlich; es macht einen substantiellen Unterschied aus, ob es um persönliche Risiken, die „Übernahme“ von Risiken für andere oder um gesellschaftliche Risiken geht.
- An Methoden zum verständigeren Umgang mit Wahrscheinlichkeiten sind zu nennen: Geeigneter Kontext mit Problemstellungen, Häufigkeitstabellen mit natürlichen Häufigkeiten, der Kalkül mit Odds.
- Wahrscheinlichkeiten, Folgekosten und Nutzen als Bausteine von stochastischen Modellen kann man kaum trennen. Gewinn-Verlust stellen sich für persönliche Risiken anders dar als für allgemeine Betrachtungen.

Die Liste mag allein wegen der Kürze als Ratgeber für Risiko unpraktisch erscheinen. Ein ganz wesentlicher Punkt ist, dass man Risiken nie isoliert beurteilen kann sondern immer nur im Vergleich von Handlungsmöglichkeiten. Viele der „Zahlen" machen eben nur dann einen Sinn. Die Informationen fehlen oft oder sind durch Interessen von Stakeholdern verzerrt. Es ist kaum möglich, sich neutrale Informationen zu beschaffen. Experten, die solche Information liefern könnten, haben in der Regel eigene Interessen. Darüber hinaus ist es zweifelhaft, eine „optimale" Lösung überhaupt anzustreben, da eine Beurteilung immer von den gewählten Kriterien abhängt; das ist anders als in der Mathematik ganz allgemein. Man erwartet von der Stochastik einfach zu viel! Ein Risikoleitfaden, der verheimlicht, dass es beste Lösungen eigentlich nicht gibt, ist nicht zielführend. Dennoch lassen wir die Leser nicht ohne Orientierung, denn es wurden zwei Kriterien durchgespielt und das beliebte Minimieren des maximalen Schadens als eine Sackgasse charakterisiert. Dagegen muss man bei einem weiter führenden Kriterium wie dem erwarteten Schaden (Gewinn etc.) größere Risiken auf sich nehmen (das Risiko von größeren Verlusten, auch wenn die Wahrscheinlichkeit dafür klein sein mag).

Stochastisches Denken ist untrennbar mit Literalität in Sachen Risiko verbunden. Wenig ist bis dato explizit formuliert, meist ist eine Schilderung von Komponenten Stochastischen Denkens an paradigmatische Beispiele geknüpft (siehe Batanero & Borovcnik, 2016; Borovcnik, 2006). Oft spielt der Begriff der bedingten Wahrscheinlichkeit eine wesentliche Rolle (Borovcnik, 2012). Für viele statistische Anwendungen ist ein tiefergreifendes Verständnis des Wahrscheinlichkeitsbegriffs eminent wichtig, weshalb an eine Aufwertung der Wahrscheinlichkeitsrechnung gedacht werden sollte, wie dies auch Borovcnik (2011) argumentiert.

Literatur

Batanero, C. & Borovcnik, M. (2016): *Statistics and probability in high school*. Rotterdam: Sense Publishers.

Borovcnik, M. (2006): Probabilistic and statistical thinking. *Proceedings of CERME 4*. Barcelona: ERME, 484-506.

Borovcnik, M. (2009): Aufgaben in der Stochastik – Chancen jenseits von Motivation. Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG) 42, 20-42.

Borovcnik, M. (2011): Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Hrsg.), *Statistics in School Mathematics.* New York: Springer (71-84).

Borovcnik, M. (2012): Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didactica de la Matemática* 1(2), 5-27. Online: www.aiem.es/index.php/aiem/.

Borovcnik, M. (2015): Risk and decision making: The "logic" of probability. *The Math Enthusiast* 12(1), 113-139.

Borovcnik, M. (2016): „To Screen or not to screen". Mathematik lehren 194, 22-28.

Borovcnik, M., & Kapadia, R. (2011a): Risk in health: more information and more uncertainty. *IASE Satellite on "Statistics Education and Outreach"*. Voorburg: ISI. Online: iase-web.org/Publications.php (6 S.).

Borovcnik, M., & Kapadia, R. (2011b): Determinants of decision-making in risky situations. *Proceedings of the 58th World Statistics Congress*. Voorburg: ISI. Online: 2011.isiproceedings.org/papers/950138.pdf (6 S.).

Borovcnik, M., & Kapadia, R. (2014): A historical and philosophical perspective on probability. In E.J. Chernoff, & B. Sriraman (Hrsg.), *Probabilistic thinking: presenting plural perspectives.* Berlin: Springer; 7-34.

Dubben, H.-H., & Beck-Bornholdt, H.-P. (2010): *Mit an Sicherheit grenzender Wahrscheinlichkeit. Logisches Denken und Zufall*. Reinbek: Rowohlt.

Gigerenzer, G. (2002): *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2007): *Gut feelings: the intelligence of the unconscious*. New York: Viking.

Kahneman, D., & Tversky, A. (1979): Prospect theory: an analysis of decision under risk. *Econom.* 47(2), 263-292.

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner & Marx.

Spiegelhalter, D. & Gage, J. (2015): What can education learn from real-world communication of risk and uncertainty? *The Mathematics Enthusiast 12*(1, 2 & 3), 4-10. Online: scholarworks.umt.edu/tme/.

Tversky, A., & Kahneman, D. (1981): The framing of decisions and the psychology of choice. *Science* 211(1), 453-458.

Verfasser

Manfred Borovcnik, Alpen-Adria-Universität Klagenfurt, Institut für Statistik, Universitätsstraße 65, 9020 Klagenfurt

manfred.borovcnik@uni-klu.ac.at