Plädoyer für "natürliche" bedingte Wahrscheinlichkeiten

HANS KILIAN, DORTMUND

Zusammenfassung: *Bedingte Wahrscheinlichkeiten werden nicht einheitlich definiert. Ich analysiere deshalb zunächst eine weniger häufig gebrauchte Variante, die mir besonders "natürlich" und deshalb auch einfach zu sein scheint. In einem weiteren Aufsatz werde ich zeigen, dass diese natürlichen bedingten Wahrscheinlichkeitsmaße auch eine spezifische Rolle in der Modellierung spielen.*

0.1 Ein Beispiel

Die folgenden Überlegungen sind durch viele Darstellungen in der Literatur, auch in Schulbüchern, angeregt worden. Ich gehe hier von einem Beispiel von A. Engel (1973; S.139,140) aus, weil ich aus seinen Büchern viel gelernt habe.

Engel, Zitat: *1. Beispiel*

Die Angehörigen eines Betriebes werden nach zwei Merkmalen eingeteilt, nach ihrem Geschlecht und nach ihren Rauchergewohnheiten.

	Frauen B	*Männer*	
Raucher A	*200*	*800*	*1000*
Nichtraucher	*300*	*200*	*500*
	500	*1000*	*1500*

Tabelle 12.1

Es sei
Ω = die Menge aller Betriebsangehörigen, A = die Menge der Raucher, B = die Menge der Frauen. Tabelle 12.1 zeigt die Zusammensetzung des Betriebs. Unser Zufallsversuch bestehe darin, eine Person zufällig auszulosen. Dadurch wird auf Ω eine Wahrscheinlichkeit P definiert, die jeder Teilmenge E die Wahrscheinlichkeit

$P(E) = \dfrac{|E|}{|\Omega|}$ *zuordnet.* (Es folgen Beispiele).

Dies sind die Anteile der Raucher, der Frauen bzw. der rauchenden Frauen im Betrieb. Wollen wir die Rauchergewohnheiten der Frauen studieren, dann ignorieren wir die Männer, und unser neuer Stichprobenraum wird die Menge B (Unterstreichungen von mir). *Unser Zufallsversuch besteht darin, eine Person aus B zufällig auszulosen. Dadurch wird auf B eine Wahrscheinlichkeit definiert, die wir P_B nennen, um sie von der alten Wahrscheinlichkeit P auf Ω zu unterscheiden. Es ist also $P = P_\Omega$. $P_B(A)$ ist die Wahrscheinlichkeit, dass eine Frau raucht. Man schreibt dafür auch $P(A|B)$. Der Anteil der Frauen unter den Rauchern ist: $P_B(A) = \ldots = \dfrac{2}{5}$*

Man nennt $P(A|B)$ die bedingte Wahrscheinlichkeit von A unter Bedingung (Hypothese) B. Die Beziehung

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)},\ \textit{(falls P(B)>0 ist) (1)}$$

wird zur Definition von $P(A|B)$ verwendet, auch für den Fall, dass kein Laplace-Experiment vorliegt. Aus (1) folgt:

$$P(A\cap B)=P(A\mid B)\cdot P(B) \qquad (2)$$

Die Beziehung (2) wird noch öfter als (1) verwendet. Man macht Annahmen über $P(A|B)$ und berechnet danach $P(A\cap B)$. (Ende des Zitates)

0.2 Was ist nun das Problem?

Das Problem wird deutlich, wenn man sich fragt, was denn die Definitionsmenge der Funktion P_B sein soll? Dabei nehme ich an, dass "Wahrscheinlichkeiten" zunächst einmal Funktionen sind, die den Ereignissen aus einer zugehörigen Definitionsmenge reelle Zahlen als deren Wahrscheinlichkeit jeweils zuordnen.

1. Antwort (A. Engel 1987, Barth-Haller 1985, Henze 2003, Kilian 1987 und viele andere): Wir verstehen unter der bedingten Wahrscheinlichkeit die Funktion

$$P(A\mid B):=\frac{P(A\cap B)}{P(B)} \text{ mit } A\in \mathcal{P}(\Omega) \qquad (0.1)$$

Dann kommt als zugehöriger Ergebnisraum (Stichprobenraum) nur ganz Ω in Frage:
$P(.\mid B): \mathcal{P}(\Omega)\to \mathbb{R}$.

Konsequenzen/Folgeprobleme: Dann gibt es in $(\Omega, P(.|B))$ im Allgemeinen viele "Ereignisse" A mit $P(A|B)=1$, nämlich alle Obermengen von B, und es gibt viele Ereignisse mit der Wahrscheinlichkeit 0, nämlich alle Teilmengen von $\Omega\backslash B$ (Ω ohne B). Also würden unter anderen das sichere Ereignis und das unmögliche Ereignis nicht mehr eindeutig dargestellt, auch nicht in endlichen Ergebnisräumen. Warum aber sollte man in dem Ergebnisraum eines endlichen Wahrscheinlichkeitsraumes Ergebnisse mit der Wahrscheinlichkeit 0 behalten, mitschleppen?

Und schließlich und auch didaktisch gesehen: Wie soll man hier Ziehen ohne Zurücklegen darstellen? So denkt dabei doch kaum jemand wirklich und so wird auch z.B. in dem schönen Buch von Barth-Haller (1989) nicht argumentiert. -

Der obige Text von Engel scheint, zunächst jedenfalls, einer anderen Antwort den Vorzug zu geben:

2. Antwort (auch Buth 2003): Wenn (B, P_B) ein (endlicher) Wahrscheinlichkeitsraum sein soll, dann muss die Potenzmenge von B (oder eine Unteralgebra von $(\mathfrak{P}(B), \cup, \cap)$) diese Definitionsmenge sein:

$$P_B : \mathfrak{P}(B) \to \mathbb{R}$$

Dann wäre allerdings in dem obigen Beispiel von Engel $P_B(A)$, A = die Menge der Raucher, gar nicht definiert, sondern man müsste statt nach dieser Größe nach $P_B(A \cap B)$ fragen. Darüber hinaus wäre auch die angegebene Definition von P(A|B) und die Multiplikationsregel

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

nicht zu verstehen, da stets $A \cap B = A$ wäre.

Dies deutet darauf hin, dass hiervon ausgehend einige Formeln und Sätze umgeschrieben werden müssten. Damit werden wir uns zunächst beschäftigen, um die Ergebnisse einer *konsequenten Durchführung dieses Ansatzes* zu klären.

1. Definitionen und Bezeichnungen

1.0 Eine Vorbemerkung

Eine Schwierigkeit im Verständnis solcher Definitionen ist, dass es im Stochastikunterricht nicht üblich ist, zwischen Variablen und Konstanten schon durch deren Benennungen deutlich zu unterscheiden. "A" ist oben an gewissen Stellen eine *Variable*, die beliebige Werte aus $\mathfrak{P}(\Omega)$ annehmen kann, an anderen Stellen ein *festes (konstantes) Ereignis*, während "B" stets ein festes Ereignis ist. Stellen Sie sich bitte kurz vor, Sie sollten Ebene Geometrie mit der allgemeinen Geradendarstellung $a \cdot b + c \cdot d + f = 0$ statt $a \cdot x + b \cdot y + c = 0$ unterrichten! Ich werde also im Folgenden zwischen Variablen und Konstanten unterscheiden.

Leider sind die Buchstaben X, Y , ... in der Wahrscheinlichkeitstheorie schon fest belegt, sonst könnte man diese Buchstaben als Bezeichner von Variablen für Ereignisse benutzen. Ich werde deshalb die Buchstaben U, V, W, ... als Bezeichner für Variable für Ereignisse bzw. für die diese darstellenden Mengen benutzen. – Gebundene Variable können weiter mit beliebigen Buchstaben bezeichnet werden.

1.1 Grundbegriffe

Es sei S eine nichtleere endliche Menge, also etwa $S = \{s_1, s_2, ..., s_n\}$ mit $n \in \mathbb{N}$ und p_S eine Funktion von S in $\mathbb{R}$: $p_S : S \to \mathbb{R}$.

Definition 1: Das Paar (S, p_S) heißt ein endlicher diskreter Wahrscheinlichkeitsraum (eDWR), wenn gilt:

(1) $p_S(s_i) \geq 0$ für alle i, und

(2) $\sum_{i=1}^{n} p_S(s_i) = 1$

Die Funktion p_S heißt die Wahrscheinlichkeitsverteilung (WV) dieses eDWR, S heißt dessen Ergebnisraum. – Bemerkung: Ich orientiere mich hier an den Bezeichnungen/Abkürzungen von Jakobs (1986).

Wir definieren nun noch das zu dem damit auch gegebenen endlichen Wahrscheinlichkeitsraum (eWR) gehörende Wahrscheinlichkeitsmaß (WM)
$P_S : \mathfrak{P}(S) \to \mathbb{R}$ mit

$$P_S(U) := \sum_{s_i \in U} p_S(s_i) \qquad (U \subseteq S) \qquad (1.1)$$

Ergänzung/Bemerkung: Insbesondere heißt die WV

$$p_S(s_i) = p_S(s_j) = \frac{1}{n} \quad \text{für} \quad i, j \in N_n = \{1, 2, 3, ..., n\}$$

eine Gleichverteilung und das dazu gehörende WM

$$P_S(U) := \frac{|U|}{|S|} \text{ für } U \in \mathfrak{P}(S) \qquad (1.2)$$

eine Laplace Wahrscheinlichkeit.

Diese Definitionen bzw. Bezeichnungen sollen auch daran erinnern, dass "Wahrscheinlichkeiten" als mathematische Objekte nicht etwas sind, dem man irgendwie in der Welt begegnet und die der Mathematiker dann aus dieser herauspräpariert, sondern als mathematische Objekte müssen sie nach unserem heutigen Verständnis ordentlich konstruiert und damit definiert und im allgemeinen auch klar bezeichnet werden. Insbesondere besteht die Konstruktion eines stochastischen Modells zumindest in der Konstruktion und das heißt Definition eines passenden WRs.

1.2 Bedingte Wahrscheinlichkeiten

Es gibt (mindestens) zwei verschiedene Definitionen bedingter Wahrscheinlichkeiten, genauer gesagt bedingter Wahrscheinlichkeitsmaße, die wir hier unterscheiden müssen.

1. Definition: Es sei (S, P_S) ein eWR und B ein Ereignis aus $\mathfrak{P}(S)$ mit $P_S(B) > 0$. Dann heißt für jedes $U \in \mathfrak{P}(S)$

$$P_S(U \mid B) := \frac{P_S(U \cap B)}{P_S(B)} \qquad (1.3)$$

die *allgemeine* bedingte Wahrscheinlichkeit von U unter der Bedingung B. Es ist leicht nachzurechnen, dass $P_S(.\,|\,B)$ ein Wahrscheinlichkeitsmaß auf S ist und also $(S, P_S(.\,|\,B))$ ein eWR.

Bemerkungen: "Allgemeine" bedingte Wahrscheinlichkeit: Meine Bezeichnung für diesen Aufsatz. -

Ein für das Verständnis des Folgenden wichtige Beobachtung ist, dass für festes B und alle $U \subseteq S$ mit $P_S(B) > 0$ gilt: $P_S(U|B) = P_S(U \cap B|B)$. -

Manche Autoren sind allerdings der Ansicht, dass der sozusagen natürliche Ergebnisraum für bedingte Wahrscheinlichkeiten unter der Bedingung B (mit $P_S(B) > 0$) die Menge B selbst ist. Dann erhält man die

2. Definition: Es sei (S, P_S) ein eWR und B ein festes Ereignis aus $\mathfrak{P}(S)$ mit $P_S(B) > 0$. Dann heißt für jedes $U \in \mathfrak{P}(B)$

$$P_{S|B}(U) := \frac{P_S(U)}{P_S(B)} \tag{1.4}$$

die *natürliche* bedingte Wahrscheinlichkeit von A unter der Bedingung B, und P_B ist ein Wahrscheinlichkeitsmaß mit der Definitionsmenge $\mathfrak{P}(B)$. Der zugehörige WR ist dann $(B, P_{S|B})$. Es ist wieder leicht nachzurechnen, dass dies ein wohldefinierter WR ist.

Bemerkung: Ich betrachte immer noch der Einfachheit halber nur endliche Ergebnismengen S, beschränke aber die Bezeichnung eDWR speziell auf die Paare vom Typ (S, p_S). Es ist dann wohl auch klar, was etwa mit $(S, p_S(.|B))$ bzw. $(B, p_{S|B})$ gemeint ist.

Bemerkung: Diese Bezeichnungen werden dadurch motiviert, dass daraus hervorgeht, dass mit $P_{S|B}$ ein bedingtes WM gegeben ist. Die spezielle Wahl dieser Darstellungsart wird dadurch gerechtfertigt, dass sie der allgemeinen Darstellung der Einschränkung einer Funktion $f: A \to D$ auf $B \subseteq A$ in etwa entspricht: $f|B : B \to D$. Die im allgemeinen hier noch notwendige neue Normierung von $P_{S|B}$ ist wohl selbstverständlich.

Auf einen Vorteil dieser Definition sei vorweg hingewiesen: Wenn p_S eine Gleichverteilung ist, dann ist auch $p_{S|B}$ ein Gleichverteilung und $P_{S|B}$ eine Laplace Wahrscheinlichkeit. (Beweis ist wohl klar.)

Beziehung zwischen $P_S(.|B)$ *und* $P_{S|B}$:

$P_{S|B}(U) = P_S(U|B)$ für alle $U \subseteq B$ (1.5)

$P_{S|B}(U \cap B) = P_S(U|B)$ für alle $U \subseteq S$ (1.6)

Aber es gilt nicht: $P_{S|B} = P_S(.\,|\,B)$.

*Bemerkung*en:

Die Definition (1.4) ist nicht nur einfacher als die übliche Definition (1.3), sondern sie macht auch deutlicher, worum es hier mathematisch geht: Man behält die relativen Wahrscheinlichkeiten $P_S(U)$ für alle $U \subseteq B$ bei, das heißt es gilt:

$$\frac{P_{S|B}(U)}{P_{S|B}(U')} = \frac{P_S(U)}{P_S(U')} \qquad \text{für alle } U, U' \subseteq B,$$

muss aber zusätzlich die Normierungsbedingung erfüllen.

Das gilt so allgemein für $P_S(.\,|\,B)$ nicht. Vielmehr ist zwar

$$\frac{P_S(U\,|\,B)}{P_S(U'\,|\,B)} = \frac{P_S(U \cap B)}{P_S(U' \cap B)} \quad \text{für alle } U, U' \subseteq S,$$

also aus dem Definitionsbereich von $P_S(.\,|\,B)$, aber

$$\frac{P_S(U\,|\,B)}{P_S(U'\,|\,B)} = \frac{P_S(U)}{P_S(U')} \quad \text{nur für } U, U' \subseteq B.$$

Es bleiben nicht *alle* relativen Wahrscheinlichkeiten erhalten!

Andererseits muss man jetzt darauf achten, welche Werte man für U in $P_{S|B}(U)$ einsetzt:

$\mathfrak{P}(B) = \mathfrak{P}(S) \cap B = \{A \cap B | A \in \mathfrak{P}(S)\}$.

Bemerkung: Für einen gegebenen WR (S, P_S) gilt im allgemeinen: Für jedes $A \subseteq S$ ist $P_S(A)$ die Wahrscheinlichkeit mit der A eintritt, wenn der zu S gehörige Versuch V_S, dessen mögliche Ergebnisse S angibt, ausgeführt wird (bei frequentistischer Interpretation von Wahrscheinlichkeitsaussagen). Dies ist aber im allgemeinen nicht der Fall für WR des Typs $(S, P_S(.\,|\,B))$. Denn für $A \subseteq S$ ist $P_S(A\,|\,B)$ die Wahrscheinlichkeit mit der das Ereignis $A \cap B$ eintritt und zwar bezogen auf diejenigen Versuchsausgänge, bei denen B auch eingetreten ist. Man kann sich vorstellen, dass zunächst ein Filter dem Versuch nachgeschaltet wird, der nur die Ausgänge durchlässt, bei denen B eingetreten ist, und dann werden nur die Ausgänge gezählt bei denen A eingetreten ist.

Als Beispiel kehren wir noch einmal zu Engel zurück: Er sagt selbst: $P_B(A)$ (= $P(A\,|\,B)$) ist die Wahrscheinlichkeit, dass eine Frau raucht. Aber das Argument der Funktion P(.|B) ist A, also das Ereignis "Ein Beschäftigter / eine Beschäftigte des Betriebes ist ein Raucher / eine Raucherin". Das ist schon eine ziemlich verwirrende Situation! Diese Darstellung stammt eben noch aus einer Zeit, als "Wahrscheinlichkeiten" eine andere inhaltliche Bedeutung hatten als nach Kolmogorov.

Ergänzung: Einer der Gutachter bzw. eine der Gutachterinnen dieses Aufsatzes hat darauf hingewiesen, dass diese

Vorstellung von bedingten Wahrscheinlichkeiten dem lernpsychologischen Modell Minerva-DM (DM = decision making) für einen sehr breiten Bereich von Wahrscheinlichkeits (Likelihood)-Aussagen von Dougherty et. al. (1999) S. 183ff entspricht. Zitat: "Erwähnenswert ist außerdem, dass der Vorschlag des Autors, dass man sich einen Filter vorstellen solle, der zuerst nur B zählt und daraus dann die A Fälle, tatsächlich psychologischen Denkmodellen zur Einschätzung von bedingten Wahrscheinlichkeiten entspricht, d.h. der Vorschlag des Autors kann sogar aus lernpsychologischen Gründen als leichter verständlich angesehen werden.

Der Autor könnte weiterhin anführen, dass seine Argumentation auch ein Grund für das leichtere Verständnis von Häufigkeiten sein könnte, da dort die Konstante (z. B.: "x von 5" !) spezifiziert wird." (Ende des Zitats)

2. Produktregeln und der Satz von der totalen Wahrscheinlichkeit für natürliche bedingte Wahrscheinlichkeiten.

2.1 Produktregeln (für natürliche bedingte Wahrscheinlichkeiten)

Gegeben sei ein eWR (S, P_S). Dann gilt für alle

(1) $A,B \in \mathfrak{P}(S)$ mit $P_S(B) > 0$:

$$P_S(A \cap B) = P_{S|B}(A \cap B) \cdot P_S(B) \qquad (1.7)$$

(2) $A,B,C \in \mathfrak{P}(S)$ mit $P_S(A \cap B) > 0$:

$$P_S(A \cap B \cap C) = P_S(A) \cdot P_{S|A}(A \cap B) \cdot P_{S|A \cap B}(A \cap B \cap C)$$

Beweis von (1): Dies ergibt sich unmittelbar aus der Definition von $P_{S|B}$

Beweis von (2):

$$P_S\big(A \cap (B \cap C)\big) = P_{S|B \cap C}\big(A \cap (B \cap C)\big) \cdot P_S(B \cap C)$$
$$= P_{S|B \cap C}\big(A \cap (B \cap C)\big) \cdot P_{S|C}(B \cap C) \cdot P_S(C)$$

Vertauscht man hier $A \leftrightarrow C$, so erhält man die Behauptung.

Zum Vergleich: Die einfachste Produktregel für allgemeine bedingte Wahrscheinlichkeiten ist bekanntlich:

Gegeben sei ein eWR (S, P_S). Dann gilt für alle $A,B \in \mathfrak{P}(S)$ mit $P_S(B) > 0$:

$$P_S(A \cap B) = P_S(A \mid B) \cdot P_S(B).$$

Die Gleichung (1.6) erlaubt es, ohne große Umrechnungen zwischen den beiden Formeltypen zu wechseln. Ich werde deshalb im Folgenden nur noch die für natürliche bedingte Wahrscheinlichkeiten angeben. -

2.2 Der Satz von der totalen Wahrscheinlichkeit für nat. bedingte Wahrscheinlichkeiten.

Unter einer Zerlegung Z einer nichtleeren Menge S in die nichtleeren Teilmengen $S_1, S_2, ..., S_m$ verstehen wir:

$S = \bigcup_{i=1}^{m} S_i$ und $S_i \cap S_j = \varnothing$ für alle $i, j \in N_m = \{1, 2, 3, ..., m\}$ mit $i \neq j$. Bezeichnung: $Z = (S_1, S_2, ..., S_m)$

Satz: Es sei (S, P_S) ein eWR und $Z = (S_1, S_2, ..., S_m)$ eine Zerlegung von S. Dann gilt für jedes Ereignis $U \in \mathfrak{P}(S)$:

$$P_S(U) = \sum_{i=1}^{m} P_{S|S_i}(U \cap S_i) P_S(S_i) \qquad (1.8)$$

Beweis: Für jedes $U \in \mathfrak{P}(S)$ gilt:

$$P_S(U) = P_S(U \cap \bigcup_{i=1}^{m} S_i)$$
$$= P_S(\bigcup_{i=1}^{m}(U \cap S_i)) = \sum_{i=1}^{m} P_S(U \cap S_i)$$

da $(U \cap S_i) \cap (U \cap S_j) = \varnothing$ für $i \neq j$. Aus der obigen Produktregel (1.7) folgt damit die Behauptung.

Bemerkungen: Ich finde, dass (2.2) leichter zu verstehen ist als die übliche Formulierung dieses Satzes und dadurch auch plausibler ist. Man sieht sozusagen, was bei einer Zerlegung von S passiert, und dass bei bedingten Wahrscheinlichkeiten immer die zugehörige Zerlegung mitgedacht werden muss, im einfachsten Fall die in B und $\overline{B}$.

Der Satz ist eine Aussage über Wahrscheinlichkeits*funktionen*, nicht nur über einzelne Wahrscheinlichkeiten, d. h. deren Werte.

Die rechte Seite der Formel dieses Satzes ist eine Linearkombination von dem Typ gewichteter Mittelwert, mit den Gewichten $P_S(S_i)$. Demgemäß sollte man eigentlich die Reihenfolge der Faktoren vertauschen:

$$P_S(U) = \sum_{i=1}^{m} P_S(S_i) P_{S|S_i}(U \cap S_i).$$

2.3 Der Satz von Bayes für natürliche bedingte Wahrscheinlichkeiten

Es seien zunächst A, B Ereignisse eines WR (S, P_S) mit $P_S(A) \cdot P_S(B) > 0$. Dann gilt nach (2.1):

$$P_{S|B}(A \cap B) \cdot P_S(B) = P_S(A \cap B)$$
$$= P_S(B \cap A) = P_{S|A}(B \cap A) \cdot P_S(A)$$

Satz von Bayes: Es sei $Z = (S_i, S_2, ..., S_m)$ wieder eine Zerlegung von S. Dann folgt (aus der obigen Formel für A=S_j):

$$P_{S|B}(S_j \cap B) := \frac{P_{S|S_j}(B \cap S_j) \cdot P_S(S_j)}{P_S(B)} \quad (1 \leq j \leq m).$$

Dies ist im Wesentlichen der Satz von Bayes für nat. bedingte Wahrscheinlichkeiten. Natürlich kann man für den Nenner auch noch einsetzen.

Bemerkung: Nach meinem Verständnis dieses Satzes kommt darin keine Variable vor: Es handelt sich nicht um einen Zusammenhang von Funktionen.

2.4 Stochastische Abhängigkeit / Unabhängigkeit in eingeschränkten Wahrscheinlichkeitsräumen

Satz: Die Ereignisse $A, B \subseteq S$ mit $P_S(A) > 0$ und $P_S(B) > 0$ sind genau dann stochastisch unabhängig, wenn gilt:

$$P_{S|B}(A \cap B) = P_S(A)$$
$$\text{bzw. } P_{S|A}(A \cap B) = P_S(B). \tag{1.9}$$

Beweis: (1) Es seien A und B stochastisch unabhängig voneinander, dh: $P_S(A \cap B) = P_S(A) \cdot P_S(B)$. Daraus folgt:

$$P_{S|B}(A \cap B) = \frac{P_S(A \cap B)}{P_S(B)} = \frac{P_S(A) \cdot P_S(B)}{P_S(B)} = P_S(A).$$

(2) Es sei nun $P_{S|B}(A \cap B) = P_S(A)$. Daraus folgt mit $P_{S|B}(A \cap B) := \frac{P_S(A \cap B)}{P_S(B)}$ die Behauptung.

Der nun noch folgende Satz hat mich zunächst sehr überrascht.

Satz: Es sei (S, P_S) ein eWR und $B \in \mathfrak{P}(S)$ ein festes Ereignis mit $P_S(B) > 0$. Dann gilt:

(1) Es seien E, F $\in \mathfrak{P}$(B) zwei Ereignisse aus (B, $P_{S|B}$), die stochastisch unabhängig sind (bzgl. (B, $P_{S|B}$)), und es sei $P_S(B) < 1$. E, F sind dann bzgl. (S, P_S) *nicht* stoch. unabhängig, sondern es gilt:

$$P_S(E \cap F) \cdot P_S(B) = P_S(E) \cdot P_S(F)$$

(2) Es seien E, F, B $\in \mathfrak{P}$(S) drei Ereignisse aus (S, P_S), die stochastisch unabhängig sind, und es sei $P_S(B) > 0$. Dann sind $E \cap B$ und $F \cap B$ auch stochastisch unabhängig bzgl. (B, $P_{S|B}$).

Bemerkung zu (1): E, F müssen eben beide in B enthalten sein, beide ziehen B nach sich.

Beweis von (1): Es gilt $P_{S|B}(E \cap F) = P_{S|B}(E) \cdot P_{S|B}(F)$, da E, F stochastisch unabhängig sind in (B, P_B). Daraus folgt aber:

$$\frac{P_S(E \cap F)}{P_S(B)} = \frac{P_S(E)}{P_S(B)} \cdot \frac{P_S(F)}{P_S(B)}$$

und daraus die Behauptung.

Beweis von (2): Wir betrachten

$$P_{S|B}((E \cap B) \cap (F \cap B)) = \frac{P_S(E \cap F \cap B)}{P_S(B)}$$

$$= \frac{P_S(E) \cdot P_S(F) \cdot P_S(B)}{P_S(B)} = P_S(E) \cdot P_S(F)$$

da E, F, B unabhängig sind. Weiter folgt nach (1.9) $= P_{S|B}(E \cap B) \cdot P_{S|B}(F \cap B)$.

Bemerkung: Man überlegt sich leicht ein Beispiel von Ereignissen E, F, bei denen E und F sowohl bzgl. (S, P_S) wie auch bzgl. (B, $P_{S|B}$) stochastisch abhängig sind, etwa weil $P_S(E \cap F) = 0$ ist, aber alle Einzelwahrscheinlichkeiten sind $\neq 0$.

2.5 Zusammenfassendes Beispiel

Um die weit reichende Bedeutung dieser Darstellung deutlich zu machen, bespreche ich noch ein einfaches Beispiel aus einem Aufsatz aus dem Jahr 2003 (Buth 2003), dessen Autor besonders sorgfältig vorgeht, so dass man auch die Stelle genau angeben kann, an der die Unstimmigkeit beginnt.

Der Autor betrachtet einen WR W=(Ω, σ, P) bestehend aus einer Menge $\Omega \neq \emptyset$, einer σ-Algebra σ über Ω und einem Wahrscheinlichkeitsmaß P, wobei klar ist $P: \sigma \to \mathbb{R}$. Dann werden bedingte Wahrscheinlichkeiten eingeführt. Sei $B \in \sigma$ mit P(B) > 0: neuer WR W_B=$(\Omega_B, \sigma_B, P_B)$ mit

$$\Omega_B := B,\ \sigma_B := \{A_B \mid A_B = A \cap B, A \in \sigma\}$$

und

$$P_B(A) := \frac{P(A \cap B)}{P(B)} \tag{1.10}$$

Damit ist eindeutig σ_B als Definitionsmenge von P_B angegeben. Aber warum dann diese Definition von P_B? Es ist doch *stets* $A \cap B = A$ für $A \in \sigma_B$. (Natürlich ist diese Definition nicht falsch.)

Als konkretes Beispiel wird der Wurf mit zwei Laplace-Würfeln betrachtet, und es werden folgende Ereignisse herausgegriffen:

A: "Die Augensumme ist höchstens 7"

B: "Die Augenzahl des ersten Würfels ist 1 oder 2"

Es ist also

$$\Omega = \{(a,b) \mid \text{mit } a,b \in N \text{ und } 1 \le a,b \le 6\}$$

$$A := \{(a,b) \in \Omega \mid a+b \le 7\}$$

$$B := \{(a,b) \in \Omega \mid a = 1 \text{ oder } a = 2\}$$

Dann wird unter anderem bestimmt

" $P_B(A) = \frac{11}{12}$ für A unter der Voraussetzung B ",

aber A ist überhaupt kein Element von σ_B. Es müsste heißen:

$P_B(A_B) = P_B(A \cap B) = \frac{11}{12}$ für A unter der Voraussetzung B.

Der Satz von der totalen Wahrscheinlichkeit wird hergeleitet in der Form

$$P(A) = P_B(A)P(B) + P_{\overline{B}}(A)P(\overline{B}) \ ,$$

aber außer $\varnothing$ gibt es überhaupt kein Ereignis, das sowohl in B wie auch in $\overline{B}$ enthalten ist. Dieser Satz muss bei diesem Aufbau heißen:

$$P(A) = P_B(A \cap B)P(B) + P_{\overline{B}}(A \cap \overline{B})P(\overline{B})$$

Alle rechnerischen Ergebnisse stimmen natürlich. Das erklärt sich so, dass die Definition (1.10) ja für allgemeine bedingte Wahrscheinlichkeiten richtig ist und alle Rechnungen damit ausgeführt werden. Es liegt nur eine Unstimmigkeit in der Struktur des Aufbaus vor.

2.6 Der Satz vom totalen Erwartungswert

Als Beispiel für die Wirkung auf Zufallsgrößen (Zufallsvariablen) möchte ich noch den Satz vom totalen Erwartungswert (Kilian 1987) behandeln:

Definitionen:

(1) Es sei (S, p_S) ein endlicher diskreter Wahrscheinlichkeitsraum und $X: S \to \mathbb{R}$ (irgend-)eine Funktion von S nach $\mathbb{R}$. Dann ist der Erwartungswert E(X) definiert als

$$E(X) := \sum_{i=1}^{n} X(s_i) \cdot p_S(s_i) \qquad (1.11)$$

(2) Es sei $B \subseteq S$ irgendeine Teilmenge von S mit $P_S(B) > 0$ und $X|B: B \to \mathbb{R}$ die Einschränkung von X auf B. Dann ist $E(X|B)$ der (bedingte) Erwartungswert von X|B, und es gilt offensichtlich:

$$E(X|B) = \sum_{s \in B} X|B(s) p_{S|B}(s) = \frac{1}{P_S(B)} \sum_{s \in B} X(s) p_S(s) \qquad (1.12)$$

Satz vom totalen Erwartungswert: Es sei nun wieder $Z = (S_1, S_2, ..., S_m)$ eine Zerlegung von S. Dann gilt:

$$E(X) = \sum_{j=1}^{m} E(X|S_j) \cdot P_S(S_j) \qquad (1.13)$$

Der Beweis ist über (1.12) fast trivial.

Bemerkung: Hier passt die oben angesprochene Bezeichnung X|B für die Einschränkung der Funktion X auf B besonders gut!

Zum Abschluss noch eine Bemerkung über *Bezeichnungen*: Bei Barth/Haller findet man die folgende Liste von Bezeichnungen für (ihr) $P_B(A)$:

Wahrscheinlichkeit des Ereignisses A

- unter der Bedingung B (ihre eigene Bezeichnung)
- unter der Voraussetzung, dass B eingetreten ist
- unter Annahme, dass B eingetreten ist
- unter der Voraussetzung, dass B eintritt
- unter der Annahme, dass B eintritt
- wenn man schon weiß, dass B eingetreten ist
- falls B
- unter der Hypothese B

Alle diese Bezeichnungen weisen eher auf Situationen hin, in denen bedingte Wahrscheinlichkeiten bei der Konstruktion geeigneter mathematischer Modelle für einen Sachverhalt verwendet werden können, als auf einen mathematischen Sachverhalt. Damit soll sich ein weiterer Aufsatz beschäftigen. Rein mathematisch gesehen handelt es sich einfach um die Einschränkung einer Struktur auf einen Teilbereich der Trägermenge, wie er überall in der Mathematik vorkommt. Buth verweist in diesem Zusammenhang auf den Begriff der "Spur" in der Topologie. Trotzdem können wir weiterhin von "bedingten Wahrscheinlichkeiten" reden.

Zusammenfassung: Diese Einführung (natürlicher) bedingter Wahrscheinlichkeiten bedeutet *strukturell* gesehen die Zerlegung eines WRs in nichtleere, disjunkte Teilmengen, im einfachsten Falle einfach in B und $\overline{B}$, auf die die alten Wahrscheinlichkeiten, nur neu normiert, direkt übertragen werden. Man muss dann allerdings einige vertraute Formeln in Definitionen und Sätzen, welche die Zusammenhänge zwischen dem Ausgangswahrscheinlichkeitsraum (S, P_S) und den Komponenten der Zerlegung, (S, $P_{S|B}$), ... beschreiben, umformulieren. Dafür gewinnen das ganze Verfahren und die Formeln an Klarheit und Durchsichtigkeit und damit auch an Verständlichkeit, vermutlich auch für Schüler und Schülerinnen.

3. Übertragung auf die Schulpraxis?

Was kann man als Lehrer / als Lehrerin tun, wenn Sie morgen mit dem Gefühl in Ihren Stochastikunterricht gehen, dass sofort etwas getan werden sollte? Was lässt sich relativ einfach umsetzten? Ich spekuliere mal, skizzenhaft.

- Sprechen Sie (fast) nie mehr von *der* Wahrscheinlichkeit! *Die* Wahrscheinlichkeit existiert genau so viel und so wenig wie *das Dreieck* oder *die* Tangente oder *die* komplexe Funktion. (Warum *fast* nie? Es gibt manchmal auch Situationen wie diese!)
- Die klassische Laplace-Definition der Wahrscheinlichkeit, also

$$P_S(A) := \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}} = \frac{|A|}{|S|}$$

kann man als Definition einer Funktion verstehen lernen, in der S eine Konstante, ein festes Ereignis ist und A eine Variable, die als "Werte" alle Elemente der Potenzmenge von S annehmen kann:

$$P_S : \mathcal{P}(S) \to [0,1] \text{ bzw. } U \to P_S(U), U \subseteq S.$$

Im Schulunterricht macht die Charakterisierung dieser Funktionen als "Wahrscheinlichkeitsmaßen" schon deshalb keinen Sinn, weil die S&S (Schüler und Schülerrinnen) noch niemals etwas von der "Maßtheorie" gehört haben werden und auch später kaum davon hören werden. Deshalb sollte man dort einfacher von "Wahrscheinlichkeits*funktionen*" reden. Zu jeder Funktion gehört eine Definitionsmenge und da es bei Wahrscheinlichkeiten üblich ist, immer "P" im Funktionsnamen zu gebrauchen, unterscheiden wir die verschiedenen Wahrscheinlichkeitsfunktionen durch den jeweiligen Zusatz "S" bzw. "B" usw..

- Bedingte Wahrscheinlichkeitsfunktionen sind Einschränkungen von Wahrscheinlichkeitsfunktionen auf eine Teilmenge B von S mit $P_S(B) > 0$. (Vielleicht erinnern sich ihre Schüler noch an die Quadratwurzelfunktion $\sqrt{\ }$ und arccos usw. als Umkehrfunktionen von durch Einschränkung entstandenen Funktionen.) Im Grunde genommen sollte man diese Funktionen auch als "*eingeschränkte* Wahrscheinlichkeitsfunktionen" bezeichnen.
- Den Satz von der totalen Wahrscheinlichkeit würde ich *für S&S* – im einfachsten Fall - so schreiben:

$$P_S(U) = P_B(U \cap B)P_S(B) + P_{\overline{B}}(U \cap \overline{B})P_S(\overline{B}) \quad (B \subseteq S)$$

Diese Darstellung geht aus der Schulbuchformel $P(U) = P(U\,|\,B)P(B) + P(U\,|\,\overline{B})P(\overline{B})$ hervor, indem man zunächst alle Funktionsnamen "P" mit dem richtigen Zusatzindex "S" bzw. "B" versieht, und dann den Strich " | " überall durch " $\cap$ " ersetzt, was keine schlechte Interpretation dieses Striches in diesem Kontext ist. Auf diese Weise kann man auch andere Schulbuchformeln neu formulieren.

Bemerkungen: Die S&S sollten aus diesem Satz lernen, dass mit jeder Einführung einer bedingten Wahrscheinlichkeitsfunktion $P_B(U)$ eine Zerlegung des Ergebnisraumes S in die Teilräume B, $\overline{B}$ verbunden ist, $S = B \cup \overline{B}$ und $B \cap \overline{B} = \varnothing$, und damit die neuen Wahrscheinlichkeitsräume (B, P_B) und $(\overline{B}, P_{\overline{B}})$ gegeben sind.

Es wird im Allgemeinen nicht wichtig sein, die Bezeichnungen $(B, P_{S|B})$ und $(\overline{B}, P_{S|\overline{B}})$ immer durchzuhalten.

- Lassen Sie Ihre S&S selbst herausfinden, was aus der Definition

$$P(A\,|\,B) := \frac{P(A \cap B)}{P(B)} \qquad (P(B) > 0)$$

wird, wenn man sie analog umschreibt und berücksichtigt, dass $A \subseteq B$ sein soll. -

Natürlich ersetzen diese Bemerkungen kein sorgfältiges Unterrichtskonzept. Das soll dieser Aufsatz aber auch gar nicht leisten.

Literatur

(1) Barth, F.; Haller R. (1989): Stochastik, Leistungskurs, München: Ehrenwirth Verlag

(2) Buth, M., (2003): Methodische Anregungen zur Behandlung der bedingten Wahrscheinlichkeit. In: Der Mathematisch naturwissenschaftliche Unterricht v.56(2003) S. 391-394

(3) Dougherty, R. P.; Charles F. Gettys; Eve E. Ogden: MINERVA-DM: A Memory Processes Model for Judgements of Likelihood. In: Psychological Review v.106 (1999) No.1. 180-209

(4) Engel, A. (1973): Wahrscheinlichkeitsrechnung und Statistik, Bd. I, Stuttgart: Ernst Klett Verlag

(5) Engel, A. (1987): Stochastik, Stuttgart: Ernst Klett Verlag

(6) Jacobs, K (1986): Mathematik für Informatiker IV / Diskrete Stochastik, Kurs der Fernuniversität Hagen

(7) Kilian, H. (1987): Bedingte Erwartungswerte im Stochastikunterricht. Stochastik in der Schule *v.*7(1987) H.3, S. 24-45