# BEDINGTE ERWARTUNGSWERTE IM STOCHASTIK-UNTERRICHT

von Hans Kilian, Hagen

## 1. Einleitung

Es ist das Ziel diese Aufsatzes zu zeigen, daß die Behandlung bedingter Erwartungswerte eine naheliegende, fruchtbare und einfach durchzuführende Erweiterung des Curriculums für den Stochastikunterricht ist. Die Aussage über die Einfachheit dieses Vorhabens gilt allerdings nur für diejenigen Schüler, die den Komplex "bedingte Wahrscheinlichkeiten" gründlich verstanden haben. Sonst muß man die Einführung bedingter Erwartungswerte mit einer Wiederholung bedingter Wahrscheinlichkeiten verknüpfen. Eine implizite Wiederholung dieses Gegenstandes ist die Behandlung bedingter Erwartungswerte allemal, und diesen Nebeneffekt kann man in Rechnung stellen, wenn man überlegt, woher man die Unterrichtszeit für das neue Thema nehmen soll.

Bedingte Erwartungswerte spielen in der Wahrscheinlichkeitstheorie und der Statistik eine wichtige Rolle. Ich verweise dazu beispielsweise auf die Regressionstheorie der elementaren Statistik. Ihre Behndlung im Unterricht würde darüberhinaus dem gelegentlich gegenüber der Stochastik erhobenen Vorwurf, daß sie im Wesentlichen nur eine Sammlung von Beispielen bzw. mathematischen Minimodellen und insofern keine "richtige" Mathematik sei, entgegenwirken. Hier werden mathematische Begriffsbildungen aus dem "normalen Stochastikunterricht" variiert und verallgemeinert. Im folgenden sollen diese Parallelen bzw. Verallgemeinerungen möglichst deutlich herausgearbeitet werden.

## 2. Bedingte Wahrscheinlichkeiten

Ich rekapituliere deshalb ganz kurz die in diesem Zusammenhang wesentlichen Aspekte bedingter Wahrscheinlichkeiten. Gegeben ist zunächst ein endlicher oder diskreter Wahrscheinlichkeitsraum $W = (\Omega;P)$ , bestehend aus einer nichtleeren Menge $\Omega$ , dem Ergebnisraum, und einer Funktion $P$ , die jeder Teilmenge von $\Omega$ , den Ereignissen, eine reelle Zahl aus dem Intervall $[0;1]$ als Wahrscheinlichkeit zuordnet und die deshalb gewissen Bedingungen genügen muß, die hier nicht aufgezählt zu werden brauchen. $P$ heißt die Wahrscheinlichkeitsbelegung von $W$ .



Dann wird, zu im Prinzip jedem $D \subseteq \Omega$ mit $P(D) \neq 0$ , eine neue Wahrscheinlichkeitsbelegung $P(.|D)$ *) zu dem alten Ergebnisraum $\Omega$ und damit ein neuer Wahrscheinlichkeitsraum $W_D := (\Omega;P(.|D))$ eingeführt. Wie dieses Vorgehen motiviert wird, sei hier ebenfalls dahingestellt, wird aber im folgenden noch eine Rolle spielen. Wichtig ist, daß zu den Wahrscheinlichkeitsräumen $W$ und $W_D$ dieselben Ergebnisräume und dieselben Ereignisalgebren, hier $\mathfrak{P}(\Omega)$ , gehören, aber verschiedene Wahrscheinlichkeitsbelegungen. Zwischen diesen besteht, per definitionem, der folgende Zusammenhang:

$$(1) \qquad P(A|D) := \frac{P(A \cap D)}{P(D)} \quad , (A \subseteq \Omega)$$

Anmerkung: Diese Art des Vorgehens hat seine guten Gründe, insbesondere den, daß man die alte Ereignisalgebra einfach weiterverwenden kann. Andererseits erscheint die Beibehaltung von $\Omega$ als Ergebnisraum aber auch oft unnatürlich, weil man häufig gerade deshalb zu der bedingten Wahrscheinlichkeit $P(.|D)$ übergeht, weil man nur noch die Ergebnisse betrachten möchte, die zu $D$ gehören. Insbesondere ist $W_D$ im allge-

---------

*) Mit $P(.|D)$ soll die bedingte Wahrscheinlichkeitsbelegung als Funktion als Ganzes bezeichnet werden, analog wie mit $P$ die Ausgangswahrscheinlichkeitsbelegung als Ganzes bezeichnet wird. Die ebenfalls anzutreffende Bezeichnung $P_D$ hat vieles für sich, ist aber aus historischen Gründen in der Literatur kaum noch anzutreffen.

meinen kein Laplaceraum mehr, wenn W selbst einer war. Die Schüler ziehen deshalb gerade im Fall von Laplaceräumen W die Wahl von D als neuem Ergebnisraum häufig intuitiv vor.

### 3. Bedingte Erwartungwerte von Zufallsgrößen

Zu jedem Wahrscheinlichkeitsraum $W = (\Omega;P)$ sind schließlich noch Zufallsgrößen $X,Y,\ldots$ eingeführt worden. Zufallsgrößen von W sind (im diskreten Fall) einfach alle Funktionen von $\Omega$ nach $\mathbb{R}$. Unter dem Erwartungswert $E(X)$ einer Zufallsgröße X versteht man (vgl.[1]) folgende Summe:

Definition 1 :

(2) $E(X) := \sum_{\omega\in\Omega} X(\omega)P(\omega)$

bzw. mit $\Omega = \{\omega_1;\omega_2;\omega_3;\ldots\}$

$E(X) := X(\omega_1)P(\omega_1) + X(\omega_2)P(\omega_2) + \ldots + X(\omega_n)P(\omega_n)$

Häufiger wird die folgende Definition gebraucht, obwohl sie meines Erachtens für den schulischen Stochastikunterricht weniger geeignet ist (vgl.[1]):

Definition 2 :

Der Erwartungswert $E(X)$ einer Zufallsgröße X eines endlichen Wahrscheinlichkeitsraumes $(\Omega;P)$ ist mit $A_i := \{\omega\in\Omega | X(\omega) = x_i\}$

(3) $E(X) := x_1\,P(A_1) + x_2\,P(A_2) + \ldots + x_n\,P(A_n)$

$= \sum_{i=1}^{k} x_i\,P(A_i) = \sum_{i=1}^{k} x_i\,P(X=x_i)$ .

Dabei sind $x_1;x_2;\ldots;x_n \in \mathbb{R}$ diejenigen Werte, die X annimmt.

Da somit jede Zufallsgröße X des Wahrscheinlichkeitsraumes $W = (\Omega;P)$ stets auch eine Zufallsgröße des Wahrscheinlichkeitsraumes $W_D = (\Omega;P(.|D))$ ist, ist damit im Prinzip auch schon erklärt, was unter dem Erwartungswert von X in bezug auf $W_D$ zu verstehen ist, und man braucht



eigentlich nur noch eine geeignete Bezeichnung, um diese Erwartungswerte, die sich auf $W_D$ und damit auf die Wahrscheinlichkeitsbelegung $P(.|D)$ beziehen, von den $E(X)$, bezogen auf W, bzw. P, zu unterscheiden. Die allgemein übliche und auch naheliegende Bezeichnung ist $E(X|D)$ .

Ich stelle diese Zusammenhänge noch einmal in einer einfachen Tabelle dar und einander gegenüber:

| Wahrscheilichkeitsraum W | Wahrscheinlichkeitsraum $W_D$ |
|---|---|
| Ergebnisraum $\Omega$ | Ergebnisraum $\Omega$ |
| Ereignisalgebra $\mathfrak{P}(\Omega)$ (bzw. $\mathfrak{A}$ ) | Ereignisalgebra $\mathfrak{P}(\Omega)$ (bzw. $\mathfrak{A}$ ) |
| Wahrscheinlichkeitsbelegung $P : \mathfrak{P}(\Omega) \to [0;1]$ (bzw. $\mathfrak{A}$ ) | Wahrscheinlichkeitsbelegung $P(\cdot|D) : \mathfrak{P}(\Omega) \to [0;1]$ (bzw. $\mathfrak{A}$ ) |
| Zufallsgrößen $X : \Omega \to \mathbb{R}$ | Zufallsgrößen $X : \Omega \to \mathbb{R}$ |
| Erwartungswerte $E(X)$ | Erwartungswerte $E(X|D)$ |

Die sogenannten bedingten Erwartungswerte $E(X|D)$ sind also nichts anderes als die normalen Erwartungswerte der Zufallsgröße X von $\Omega$, bezogen auf den Wahrscheinlichkeitsraum $W_D = (\Omega;P(.|D))$ . Insbesondere gilt also:

$$E(X|D) = \sum_{\omega\in\Omega} X(\omega)P(\omega|D)$$

usw. Anmerkung: Dies gilt so allerdings nur für diskrete Wahrscheinlichkeitsräume.

## 4. Eigenschaften von bedingten Erwartungswerten

Die obige Entwicklung und Erläuterung der Bezeichnung "bedingte Erwartungswerte" sollte so deutlich wie möglich machen, daß hier eigentlich nichts zu definieren ist, kein neuer Begriff gebildet wird, sondern eben nur eine neue Bezeichnung für eine spezielle Sorte von Erwartungswerten einzuführen ist. Das "eigentlich" in dieser Aussage bezieht sich darauf, daß von den mir bekannten Lehrbüchern der elementaren Wahrscheinlichkeitstheorie kaum eines diesen Sachverhalt klar herausstellt: Es gibt eigentlich nichts Neues zu definieren, und damit auch nichts zu verstehen, sondern es muß lediglich eine neue Bezeichnung eingeführt werden, und man muß sich dann vielleicht noch überlegen, wie die bedingten Erwartungswerte zu interpretieren sind.

Insbesondere haben bedingte Erwartungswerte alle die üblichen Eigenschaften von Erwartungswerten. Sie sind in der folgenden Liste zusammengestellt, zusammen mit üblichen Darstellungen von Erwartungswerten, spezialisiert auf den vorliegenden Fall. Es gibt auch hier nichts mehr zu beweisen!

Folgerung: Es seien $X$, $Y$ Zufallsgrößen eines Wahrscheinlichkeitsraumes $(\Omega;P)$ und $D$ ein Ereignis von mit $P(D) \neq 0$. Es sei $W_D = (\Omega;P(.|D))$ der zu $D$ gehörige Wahrscheinlichkeitsraum mit der bedingten Wahrscheinlichkeitsbelegung $P(.|D)$. Dann gilt für den bedingten Erwartungswert in bezug auf $W_D$ stets:

(4) $$E(X|D) = \sum_{\omega\in\Omega} X(\omega)\, P(\omega|D)$$

(5) $$E(X|D) = \sum_{i=1}^{m} x_i\, P(X=x_i|D)\,,$$

falls $x_1;x_2;\ldots;x_m$ die verschiedenen Werte sind, die $X$ annehmen kann

(6) $$E(aX|D) = a\,E(X|D) \quad (a\in\mathbb{R})$$



(7) $$E(X+Y|D) = E(X|D) + E(Y|D)$$

Etwas Neues enthält die folgende Aussage, die auch die Ausgangsbasis für die Interpretation von bedingten Erwartungswerten bildet ebenso wie für einen wichtigen Anwendungsaspekt:

Satz 1 :

(8) $$E(X|D) = \frac{1}{P(D)} \sum_{\omega\in D} X(\omega)\, P(\omega)$$

Beweis:

$$P(\omega|D) = \frac{P(\{\omega\}\cap D)}{P(D)} = \begin{cases} P(\{\omega\})/P(D)\,, & \text{falls } \omega\in D \\ 0\,, & \text{falls } \omega\notin D \end{cases}$$

Daraus ergibt sich die Behauptung mit der obigen Darstellung (4) von $E(X|D)$. QED

Im folgenden Abschnitt werden noch weitere besondere Eigenschaften bedingter Erwartungswerte angegeben werden.

## 5. Interpretation und erste Anwendungen bedingter Erwartungswerte

Interpretation und Anwendung bedingter Erwartungswerte ergeben sich aus der Interpretation und Anwendung bedingter Wahrscheinlichkeiten. Im Unterricht wird dadurch auch hier die entsprechende Wiederholung herausgefordert.

Erstens: Der Bezug auf die Bedingung, die dadurch gegeben ist, daß nur noch Versuchsausgänge/Ergebnisse betrachtet werden, bei denen das Ereignis $D$ eingetreten ist, kommt in der Darstellung (8) des bedingten Erwartungswertes $E(X|D)$ klar zum Ausdruck, besonders für den, der sich an den Beweis von (8) erinnert. Der Faktor $1/P(D)$ ist offensichtlich ein Normierungsfaktor: Es wird gewissermaßen nicht mehr der Mittelwert aus allen Meßwerten gebildet, sondern nur noch aus denen, die die Bedingung $D$ erfüllen.

Beispiel 1: Wurf mit 2 Würfeln (vgl. das Beispiel 3 in [1]).

Es sei $Z$ die Augensumme der Würfel und $G$ das Ereignis : "Die Augensumme ist gerade". Gesucht sei $E(Z|G)$ !

Lösung: Nach Satz 1 gilt

$$E(Z|G) = \frac{1}{P(G)} \sum_{\omega \in G} Z(\omega)\ P(\omega)$$

Die hier anstehende Summe berechnen wir mit Hilfe eines Ergebnisbaumes zu unserem Versuch (siehe Bild 1, vgl. [1], [3]).

Daraus folgt mit $P(G) = 1/2$ :

$$E(Z\ G) = (\frac{1}{2})^{-1} \cdot \frac{1}{36} \cdot (1\cdot 2 + 3\cdot 4 + 5\cdot 6 + 5\cdot 8 + 3\cdot 10 + 1\cdot 12)$$

$$= \frac{2}{36}\ (\ 1\cdot(2+12) + 3\cdot(4+10) + 5\cdot(6+8))$$

$$= \frac{1}{18}\cdot 14\cdot(1+3+5) = 7$$

Zweitens: Die wichtigsten elemetaren Anwendungen bedingter Wahrscheinlichkeiten erfolgen vermutlich über den Satz von der totalen Wahrscheinlichkeit: Es sei $D_1;D_2;\ldots;D_m$ eine Zerlegung des Ergebnisraumes $\Omega$ mit $P(D_i) \neq 0$ für $1 \leq i \leq m$ . Dann gilt bekanntlich für jedes Ereignis $A$ :

$$(9)\quad P(A) = P(A|D_1)\ P(D_1) + P(A|D_2)\ P(D_2) + \ldots + P(A|D_m)\ P(D_m)$$

Dieser Satz wird häufig angewandt, um die Wahrscheinlichkeiten $P(A)$ zu berechnen oder zu definieren, da sich häufig die Wahrscheinlichkeiten $P(A|D_i)$ leichter angeben lassen, da man dabei über die zusätzliche Information verfügt, daß $D_i$ eingetreten ist. Im Grunde handelt es sich um die Einführung einer (vollständigen) Fallunterscheidung, und diese



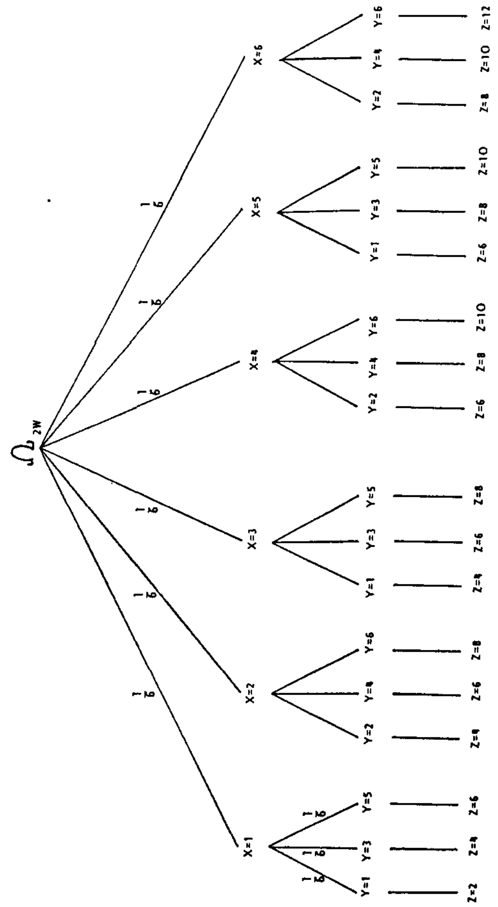

Bild 1

allgemeine Überlegung läßt uns vielleicht vermuten, daß es eine analoge Aussage auch für Erwartungswerte geben muß. Dieses ist der Satz über den totalen Erwartungswert:

Satz 2 :

Es sei $D_1;D_2;\ldots;D_m$ eine Zerlegung von $\Omega$ mit $P(D_j) \neq 0$ für jedes $j \in \{1;2;\ldots;m\}$. Dann gilt für jede Zufallsgröße $X$ von $W = (\Omega;P)$ bzw. $W_D = (\Omega;P(.|D)$:

$$(10) \quad E(X) = E(X|D_1)\,P(D_1) + E(X|D_2)\,P(D_2) + \ldots + E(X|D_m)\,P(D_m)$$

**1.Beweis :** Der übliche Beweis beruht im wesentlichen auf einer Anwendung des Satzes über die totale Wahrscheinlichkeit. Nach (5) gilt mit $X(\Omega) = \{x_1;x_2;\ldots;x_k\}$ :

$$E(X) = \sum_{i=1}^{k} x_i\,P(X=x_i)$$

$$= \sum_{i=1}^{k} x_i \sum_{j=1}^{m} P(X=x_i|D_j)\,P(D_j)$$

nach (9), angewandt auf $P(X=x_i)$ für jedes $i \in \{1;2;\ldots;k\}$. Vertauscht man die Reihenfolge der Additionen, so ergibt sich

$$E(X) = \sum_{j=1}^{m} P(D_j) \sum_{i=1}^{k} x_i\,P(X=x_i|D_j)$$

Der zweite Faktor in der j-Summe ist nach (5) gerade der bedingte Erwartungswert $E(X\,D_j)$ und also folgt:

$$E(X) = \sum_{j=1}^{m} E(X|D_j)\,P(D_j)$$

QED

Dieser Beweis dürfte den Schülern i.a. kaum zugänglich sein. Von der in [1] begründeten Beobachtung, daß die



Definition 1 des Erwartungswertes für theoretische Überlegungen besonders gut geeignet ist, versuchen wir deshalb einen Beweis über diese Darstellung von Erwartungswerten.

**2.Beweis :** Wir gehen aus von der Summe

$$E(X|D_1)\,P(D_1) + E(X|D_2)\,P(D_2) + \ldots + E(X|D_m)\,P(D_m)\ ,$$

weil es meistens einfacher ist, einen komplizierten Ausdruck zu vereinfachen, als umgekehrt und erhalten nach Satz 1

$$= P(D_1)\,\frac{1}{P(D_1)} \sum_{\omega \in D_1} X(\omega)P(\omega) + P(D_2)\,\frac{1}{P(D_2)} \sum_{\omega \in D_2} X(\omega)P(\omega) + \ldots + P(D_m)\,\frac{1}{P(D_m)} \sum_{\omega \in D_m} X(\omega)P(\omega)$$

nach der Darstellung (4), angewandt auf jedes $E(X|D_i)$ . Daraus folgt offensichtlich

$$= \sum_{\omega \in D_1} X(\omega)P(\omega) + \sum_{\omega \in D_2} X(\omega)P(\omega) + \ldots + \sum_{\omega \in D_m} X(\omega)P(\omega)$$

Da nun jedes $\omega \in \Omega$ in genau einem der $D_i$ vorkommt, erhält man als gesamte Summe einfach (vgl. 1 )

$$= \sum_{\omega \in \Omega} X(\omega)\,P(\omega) = E(X)$$

QED

Dieser Beweis ist nicht nur einfacher als der 1.Beweis, sondern er verschafft uns auch mehr Einsicht in den Sachverhalt: Es wird sehr deutlich, wie die Zerlegung $D_1;\ldots;D_m$ benutzt wird, um die verschiedenen Fälle auseinander zu sortieren, und wie die "Gewichte" $P(D_i)$ für die richtige Normierung der einzelnen bedingten Erwartungswerte sorgen.

Ein besonders wichtiger Speialfall von (10) ist der, daß die Zerlegung von $\Omega$ dadurch gegeben ist, daß eine zweite Zufallsgröße $Y$ die verschiedenen Werte $y_1;y_2;\ldots;y_s$

annimmt. Dann folgt:

$$(11)\qquad E(X) = \sum_{j=1}^{s} E(X|Y=y_j)P(Y=y_j)$$

**Beispiel 2** : Als erstes Beispiel zum totalen Erwartungswert betrachten wir den Erwartungswert des Produktes Pr beim Wurf mit 2 Würfeln. Es sei $X$ die Augenzahl des roten Würfels und $Y$ die des schwarzen Würfels. Dann folgt mit (11) :

$$E(Pr) = E(XY) = \sum_{j=1}^{6} E(XY \; Y=j)P(Y=j)$$

$$= \frac{1}{6} \sum_{j=1}^{6} E(jX|Y=j) = \frac{1}{6} \sum_{j=1}^{6} j\, E(X|Y=j)$$

mit $E(x|Y=j) = 7/2$

$$= \frac{1}{6}\cdot\frac{7}{2}\cdot \sum_{j=1}^{6} j = \frac{1}{6}\cdot\frac{7}{2}\cdot\frac{6\cdot 7}{2} = 12.25$$

Die hier benutzte Aussage

$$E(XY|Y=j) = E(jX|Y=j) = jE(X|Y=j)$$

würde ich im Unterricht als intuitiv klar betrachten. Man kann sie aber auch allgemein beweisen (siehe unten). Die weitere Aussage $E(X|Y = j) = E(X)$ würde ich in diesem Würfelbeispiel ebenfalls als klar betrachten. Man kann das aber auch nachrechnen, ausgehend von

$$\{Y = j\} = \{(1;j);(2;j);\ldots;(6;j)\} ,$$

oder einfach an einem Ergebnisbaum ablesen (siehe Bild 2).



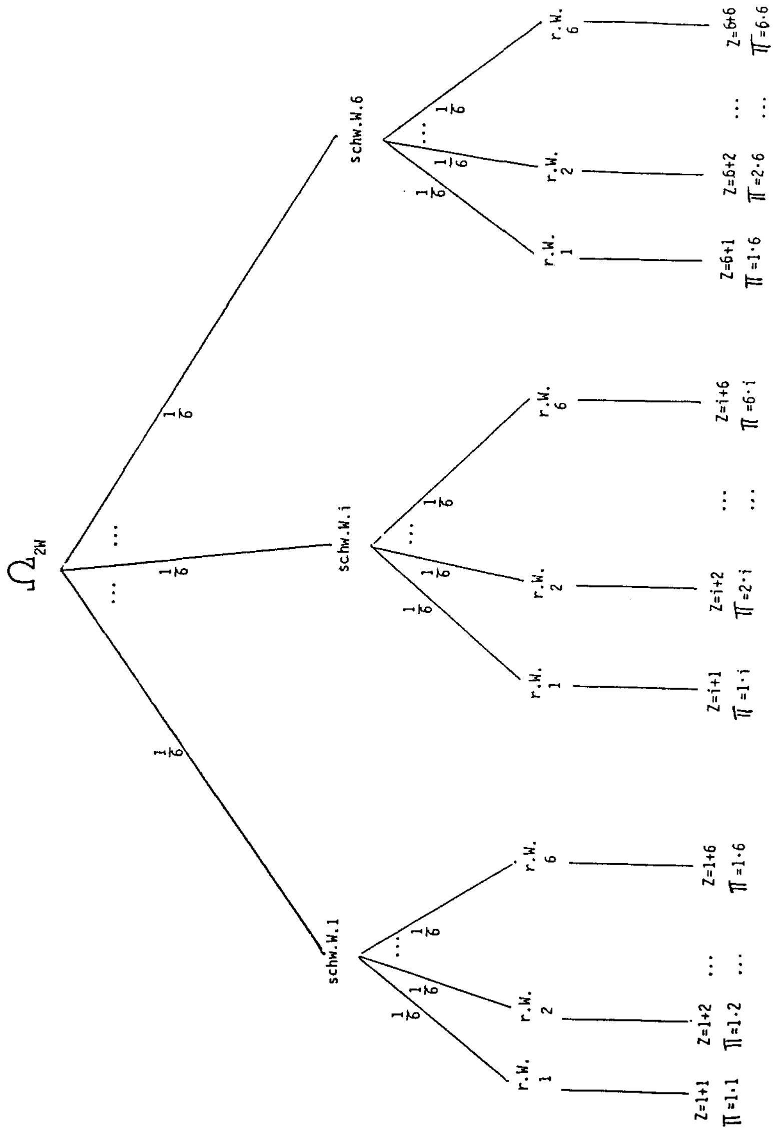


Bild 2

Der folgende Satz 3 faßt noch einige wichtige Eigenschaften von bedingten Erwartungswerten in endlichen (bzw. diskreten Wahrscheinlichkeitsräumen zusammen.

Satz 3 :

Es seien $X,Y$ Zufallsgrößen eines (diskreten Wahrscheinlichkeitsraumes $W = (\Omega;P)$ , $g$ irgendeine reelwertige Funktion und $y \in Y(\Omega)$ . Dann gilt stets:

(12) I $E(X|\Omega) = E(X)$

(13) II $E(X \cdot g(Y)|Y=y) = g(y) \cdot E(X|Y=y)$

III $E(X+g(Y)|Y=y) = E(X|Y=y) + g(y)$

Bemerkung: Oft definiert man noch $E(X|D) = 0$ , falls $P(D) = 0$ ist. Dann kann man bei dem Satz über den totalen Erwartungswert die Voraussetzung weglassen, daß alle $P(D_i) \neq 0$ sein müssen, und man kann I ergänzen um $E(X|\emptyset) = 0$ .

Beweis von II : Nach Satz 1 gilt:

$$E(x \cdot g(Y)|Y=y) = \frac{1}{P(Y=y)} \sum_{\omega \in \{Y=y\}} X(\omega) \cdot g(Y(\omega)) \cdot P(\omega)$$

In dieser Summe kommen nur solche Summanden vor, in denen $Y(\omega)$ den Wert $g(y)$ hat. Also kann man $g(y)$ ausklammern und erhält

$$E(x \cdot g(Y)|Y=y) = g(y) \frac{1}{P(Y=y)} \sum_{\omega \in \{Y=y\}} X(\omega) P(\omega)$$

$$= g(y) E(X|Y=y)$$

unter nochmaliger Anwendung von Satz 1.

Ich erwähne hier nur, daß für unabhängige Zufallsgrößen $X,Y$ und $y \in Y(\Omega)$ stets $E(X|Y=y) = E(X)$ ist. Der Beweis ist leicht, wenn man von der Darstellung (5) ausgeht.



## 6. Vermischte Beispiele

Es folgt noch eine Reihe von Beispielen, hauptsächlich zur Berechnung von totalen Erwartungswerten. Für mehrere von ihnen ist es charakteristisch, daß ähnlich wie in dem Würfelspiel gewisse bedingte Erwartungswerte als aus der Situation heraus intuitiv klar behandelt werden. Auf diese Weise kann man auch in der Schule zu die Schüler überzeugenden Lösungen von Problemen kommen, die ihnen sonst gar nicht oder nur über eine längere Rechnung zugänglich wären.

Beispiel 3 : Das Problem des Diebes von Bagdad

The thief of Bagdad has been placed in a dungeon with three doors. One of the doors leads into a tunnel that returns him to the dungeon after one day's travel through the tunnel. Another door leads to a similar tunnel whose traversal requires three days rather than one day. The third door leads to freedom.

Assume that the thief is equally likely to choose each door each time he makes a choice. (The thief has a poor memory!) Find the exspected number of days the thief will be imprisoned, from the moment he first chooses a door to the moment he chooses the door leading to freedom. Assume that the thief immidiately chooses a door each time he enters the dungeon.

dungeon: Kerker,Verlies ; traversal: Durchgang,Durchquerung

Lösung: Wir wenden den Satz vom totalen Erwartungswert in der Form eines Ereignisbaumes an. Es sei T die Dauer bis zur Wahl der richtigen Tür. Dann gilt "offensichtlich":

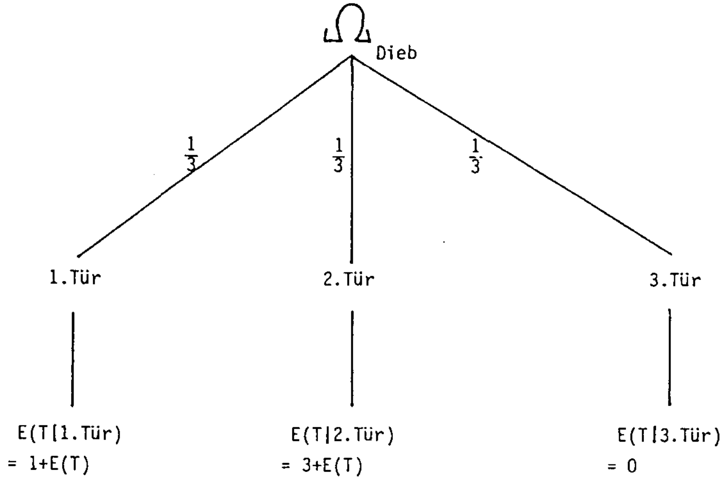


Bild 3

Denn wenn der Dieb die Tür wählt, die ihn nach einem Tag wieder in den Kerker zurückführt, so beginnt - nach den Bdingungen der Aufgabenstellung - das ganze Verfahren von neuem ! Also folgt:

$$E(T) = \frac{1}{3}\cdot(1+E(T)) + \frac{1}{3}\cdot(3+E(T)) + \frac{1}{3}\cdot 0$$

Daraus folgt $3E(T) = 4 + 2E(T)$ , also $E(T) = 4$.

**Beispiel 4** : Der Erwartungswert einer binomial verteilten Zufallsgröße

Die Größe $S_n$ sei $b(n;p)$ verteilt. Sn kann interpretiert werden als die Anzahl des Eintretens eines bestimmten Ereignisses bei $n$ unabhängigen Versuchen. Es sei $E$ das Ereignis, daß bei der ersten Durchführung des Versuches das betrachtete Ereignis eintritt, $P(E) = p$ . Dann gilt:

$$E(S_n) = E(S_n|E)\ P(E) + E(S_n|\bar{E})\ P(\bar{E})$$



$$= p\cdot E(S_n|E) + (1-p)\cdot E(S_n|\bar{E})$$

Andererseits ist intuitiv klar:

$$E(S_n|E) = 1 + E(S_{n-1}) \; ; \; E(S_n|\bar{E}) = E(S_{n-1})$$

Damit folgt insgesamt:

$$E(S_n) = p\cdot(1+E(S_{n-1})) + (1-p)\cdot E(S_{n-1})$$
$$= p + E(S_{n-1})$$

Dies ist eine Rekursionsformel zur Berechnung von $E(S_n)$ . Mit $E(S_1) = p$ folgt $E(S_n) = np$ .

**Beispiel 5** : Das obige Beispiel kann man auch auf anderen Wegen in der Schule berechnen. Das folgende dagegen bekanntlich kaum: Es sei $T$ die Wartezeit bis zum ersten Eintreten eines Ereignisses $A$ , das bei jeder Versuchsdurchführung mit der Wahrscheinlichkeit $p$ eintritt. Die Versuche sollen unabhängig voneinander durchgeführt werden. Konkretes Beispiel: Wie oft muß man im Schnitt würfeln, bis zum ersten Mal eine Sechs fällt ?

Bild 4 zeigt einen Ereignisbaum zu diesem Problem:

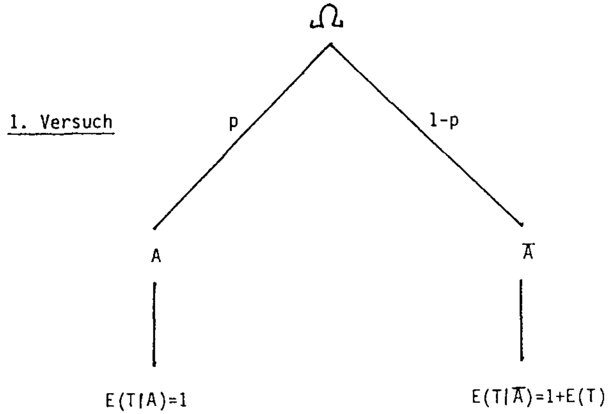


Bild 4

Damit folgt:

$$E(T) = p + (1-p)(1+E(T))$$

$$= p + (1-p) + (1-p)E(T)$$

Es folgt $p \cdot E(T) = 1$ und also $E(T) = 1/p$ . Das ist m.E. eine sehr bemerkenswerte Methode, dieses wichtige Resultat zu erreichen.



**Beispiel 6:** Stichproben aus einer Urne mit zufälliger Zusammensetzung

Aus einer Urne mit $N$ Kugeln wird eine Stichprobe vom Umfang $n$ ohne Zurücklegen gezogen. Die Anzahl $X$ der weißen Kugeln in der Urne sei eine zufällige Größe. Es sei $Y$ die Anzahl der Kugeln in der Stichprobe.

Dann gilt bekanntlich:

$$P(Y=y|X=x) = \frac{\binom{x}{y}\binom{N-x}{n-y}}{\binom{N}{n}} ,$$

und daraus folgt bekanntlich (das sei hier alles vorausgesetzt)

$$E(Y|X=x) = n\frac{x}{N}$$

Nach dem Satz über den totalen Erwartungswert erhalten wir für $E(Y)$ :

$$E(Y) = \sum_{x=0}^{N} E(Y|X=x)\cdot P(X=x)$$

$$= \frac{n}{N}\sum_{x=0}^{N} x\cdot P(X=x) = \frac{n}{N}E(X)$$

Man erhält dieses Ergebnis, ohne etwas über die Verteilung der Zufallsgröße $X$ zu wissen ! $E(X)/N$ ist der mittlere Anteil der weißen Kugeln in der Urne. Man kann also $n\cdot E(X)/N$ in Analogie zum Erwartungswert einer binomial oder einer hypergeometrisch verteilten Zufallsgröße sehen (mit $p = E(X)/N$ ).

**Beispiel 7 :** Ich verweise hier nur auf einen Aufsatz von A.F.BISSEL in dieser Zeitschrift [2] über ein Problem mit einer Nylonfadenmaschine. Dort werden Zusammenhänge zwischen Erwartungswerten angegeben, die man m.E. nur über den Satz über den totalen Erwartungswert verstehen und herleiten kann.

**Beispiel 8** : Bei einem Hersteller von Elektrogeräten treffen täglich Lieferungen eines gewissen Bauteiles ein, die einen in gewissen Grenzen zufällig schwankenden Umfang haben. Es sei $N$ die Anzahl der Bauteile pro Lieferung. Die Wahrscheinlichkeitsfunktion der Zufallsgröße $N$ sei:

| n: | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|
| P(N=n): | 0.05 | 0.10 | 0.10 | 0.20 | 0.35 | 0.20 |

Von den gelieferten Bauteilen können einzelne defekt sein. Die Wahrscheinlichkeit dafür sei für alle gelieferten Bauteile gleich groß, und zwar $p = 0.1$ . Mit wievielen defekten Bauteilen pro Lieferung muß der Empfänger im Schnitt rechnen ?

Lösung: Es sei $X$ die Anzahl der täglich ankommenden defekten Bauteile. Gesucht ist der Erwartungswert von $X$ . Nach dem Satz über den totalen Erwartungswert gilt

$$E(X) = \sum_{n=10}^{15} E(X|N=n)\cdot P(N=n)$$

Für gegebenes $n$ ist die Zufallsgröße $X$ binomial verteilt, also ist $E(X|N=n) = np = 0.1n$ .

Daraus folgt:

$$E(X) = 0.1 \sum_{n=10}^{15} n\cdot P(N=n) = 0.1\, E(N) = 1.33$$

**Beispiel 9** : Die Kosten für die Durchführung eines bestimmten Versuches mögen 10DM betragen. Der Versuch wird solange wiederholt, bis ein gewünschtes Ereignis $A$ eingetreten ist. Die Wahrscheinlichkeit für das Eintreten von $A$ bei einer einzelnen Versuchsdurchführung sei $p$ und die Wiederholungen unabhängig voneinander. Was wird eines solche Versuchsdurchführung im Mittel kosten ?



Lösung: Es sei $K$ die Zufallsgröße, die die Kosten pro Versuchsreihe angibt. Dann ist $E(K) = 10DM/p$ .

Man kann dieses Beispiel auf das Beispiel 5 zurückführen ($K = 10DM\cdot T$) oder völlig analog zu 5 direkt bestimmen.

**Beispiel 10** : Die ersten 5 Wiederholungen eines Experimentes kosten je 20DM, alle weiteren Wiederholungen kosten jeweils 10DM. Das Experiment wird solange wiederholt, bis ein Erfolg (ein bestimmtes Ereignis A) erreicht wird. Die Wahrscheinlichkeit für das Eintreten des Ereignisses $A$ bei einer einzelnen Durchführung des Experimentes sei $p$ (z.B. $p = 0.9$) und die Wiederholungen seien unabhängig voneienander.

Was wird eine solche Versuchsreihe im Mittel kosten ?

Lösung: Es sei $K$ = "Kosten einer Versuchsreihe" und $F$ das Ereignis, daß das gewünschte Ereignis $A$ spätestens bis zum fünften Experiment eingetreten ist. Dann erhalten wir folgenden Ereignisbaum (mit $q = 1-p$) zu der Fallunterscheidung $F$, $\overline{F}$ :

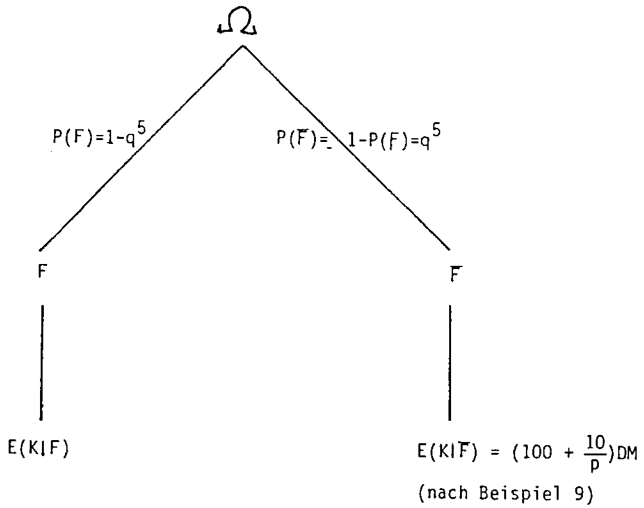


Bild 5

Es bleibt noch $E(K|F)$ zu berechnen. Nach Satz 1 erhalten wir dafür

$$E(K\ F) = \frac{1}{P(F)} (20DM\ p + 40DM\ pq + \ldots + 100DM\ pq^4)$$

$$= \frac{1}{P(F)} (1 + 2q + 3q^2 + 4q^3 + 5q^4)\ 20DM$$

Insgesamt erhalten wir also:

$$E(K) = p(1+2q+3q^2+4q^3+5q^4)20DM + q^5(100 + \frac{10}{p})DM$$

$$= (\ 20p\ (1+2q+3q^2+4q^3+5q^4) + q^5(100 + \frac{10}{p}))DM$$



Dieser Aufsatz enthält nichts grundsätzlich Neues, mit der eventuellen Ausnahme des 2. Beweises zu dem Satz 2 und der Benutzung von Ereignisbäumen. Auch die meisten Beispiele habe ich in der Literatur gefunden, und einige von ihnen finden sich in vielen Büchern über Wahrscheinlichkeitsrechnung. Deshalb erscheint es mir nicht sinnvoll, im einzelnen Literaturhinweise zu geben.

**LITERATUR**

[1] KILIAN,H.: Zur Definition von Erwartungswerten und von bedingten Erwartungswerten und zu deren Berechnung mit Hilfe von Bäumen .- Beiträge zum Mathematikunterricht 1987,S.188-191

[2] BISSEL,A.F.: Anwendungen der Wahrscheinlichkeitstheorie in der Industrie .- Stochastik in der Schule 1(1979), Heft 2, S.15-17

[3] KOOPS,H. et al.: Mathematik für Lehrer in der Sekundarstufe I / Hauptschule .- Ein Fernstudienlehrgang, HE12 Wahrscheinlichkeitsrechnung; DIFF, Tübingen 1981