Weg führte zwar ebenfalls zum Ziel, bedurfte aber großer Anstrengungen; deshalb wird an dieser Stelle auf einen Abdruck dieses relativ umfangreichen Beweises verzichtet.

Stattdessen möchte ich noch auf eine Idee in (Treiber 1988) hinweisen. Dort wird verdeutlicht, wie man die Formel (6) auf das allgemeinere *Zählprinzip des Ein- und Ausschließens* zurückführen kann.

Man kann nun die Formel des Ein- und Ausschließens mit vollständiger Induktion beweisen (Treiber verzichtet darauf, man findet den Beweis aber z. B. in (Henze 1997) auf S. 72) oder sie wieder auf ein allgemeineres Zählprinzip, den binomischen Lehrsatz zurückführen, wie es z. B. in (Danckwerts/Vogel/Bovermann 1985) auf den Seiten 128/129 und 152 auch für LK-Schüler verständlich dargestellt ist.

Literatur:

Althoff, H. (1999): Das Lösen stochastischer Probleme mit Hilfe eines Fächerbelegungsmodells – ein Bericht über eine Unterrichtsreihe in einem Leistungskurs 13. - In: Berichte aus dem Seminar für Didaktik der Mathematik WS 97/98 und SS 98, Universität Bielefeld

Danckwerts, R./Vogel, D./Bovermann, K. (1985): Elementare Methoden der Kombinatorik. - Verlag Teubner Stuttgart.

Fricke, A. (1984): Das stochastische Problem der vollständigen Serie. - In: Der Mathematikunterricht 30, Heft 1.

Henze, N. (1997): Stochastik für Einsteiger. - Verlag Vieweg Braunschweig/Wiesbaden

Strick, H. K.(1994): Modellieren als Aufgabe des Stochastikunterrichts. - In: Praxis der Mathematik 36, Heft 2.

Treiber, D. (1988): Zur Wartezeit auf eine vollständige Serie. - In: Didaktik der Mathematik 16, Heft 3.

Bedingte Erwartungswerte

Hans Humenberger, Wien

Zusammenfassung: Das Thema *Bedingte Erwartungswerte* ist in der didaktischen Literatur nicht oft zu finden. Es ist daher kein Wunder, daß sie in vielen Stochastikkursen für Schüler (und bisweilen vielleicht auch für Studenten) fehlen. Dies ist u.E. insbesondere dann „schade", wenn in einem Stochastikkurs sowohl bedingte Wahrscheinlichkeiten als auch Erwartungswerte besprochen werden. Die dann noch fehlende Symbiose zum bedingten Erwartungswert ist nicht mehr mit viel zusätzlichem Aufwand verbunden, öffnet aber eine Reihe von Einsatz- bzw. Anwendungsmöglichkeiten (siehe z.B. KILIAN 1987 oder HUMENBERGER 1998abc).

Nach einer kurzen Replik von bedingten Wahrscheinlichkeiten und Erwartungswerten bei diskreten Zufallsvariablen wird ein intuitiver Zugang zu bedingten Erwartungswerten (in Analogie zu bedingten Wahrscheinlichkeiten) kurz dargestellt. Die (etwas abstraktere) Definition eines *bedingten Erwartungswertes* ist jedoch mit dieser der Intuition entsprechenden Sichtweise verträglich. Es folgen in dieser Arbeit dann einige kurze Beispiele, die das „Potential" von bedingten Erwartungswerten illustrieren sollen.

1 Bedingte Erwartungswerte

1.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte

Bedingte Erwartungswerte stehen zu normalen Erwartungswerten im selben Verhältnis wie bedingte Wahrscheinlichkeiten zu normalen Wahrscheinlichkeiten. Deshalb sei kurz an bedingte Wahrscheinlichkeiten erinnert. Meist geschieht der Zugang zu diesen intuitiv anhand eines Beispiels wie dem folgenden. Jemand hat aus einer Personengruppe, die sich aus Damen – Herren bzw. Brillenträgern – Nicht-Brillenträgern (wie in Tab. 1 zu lesen) zusammensetzt, eine Person zufällig ausgewählt.

	D	H
B	2	3
$\neg B$	5	4

Tab. 1: Damen (D) – Herren (H), Brillenträger (B) – Nicht-Brillenträger ($\neg B$)

Man erhält aus dieser Tabelle sofort für die einzelnen Wahrscheinlichkeiten:

$$P(D) = \frac{7}{14} = \frac{1}{2}, \quad P(H) = \frac{7}{14} = \frac{1}{2},$$

$$P(B) = \frac{5}{14}, \quad P(\neg B) = \frac{9}{14}.$$

Jemand verrät einem, daß es sich bei der gewählten Person um eine Dame handelt. Wie groß ist dann die Wahrscheinlichkeit, daß diese Dame eine Brillenträgerin ist?

Auch wenn jemand noch nichts von *bedingten Wahrscheinlichkeiten* gehört hat, so wird er argumentieren: Aufgrund der Information, die ich habe, kommen nur mehr noch die 7 Damen in Frage („Anzahl der möglichen Fälle"), wobei 2 von diesen („günstige Fälle") Brillenträgerinnen sind. Der

H. Althoff: Zur Berechnung der Wahrscheinlichkeit für das Vorliegen einer vollständigen Serie (Sammelbildproblem)
H. Humenberger: Bedingte Erwartungswerte

ursprüngliche Ereignisraum Ω einer Zufallswahl (nämlich die Gesamtheit aller 14 Personen) wurde durch die Information („Bedingung") *Dame* auf 7 Personen eingeschränkt (analog wäre es, wenn man wüßte, daß es sich bei der gewählten Person um einen Herrn handelt).

Man sieht also auch intuitiv, daß der Anteil „Brillenträger unter den Damen bzw. Herren" sich durch folgende Werte charakterisieren läßt:

$$P(B|D) = \frac{2}{7},\ P(B|H) = \frac{3}{7}.$$

Diesem intuitiven Zugang folgt dann meist eine Exaktifizierung (genaue Definition) und eine Zusammenstellung bzw. Begründung gewisser Rechenregeln bei bedingten Wahrscheinlichkeiten.

Definition: Wenn A ein Ereignis mit positiver Wahrscheinlichkeit ist, so definiert man die *bedingte Wahrscheinlichkeit eines Ereignisses B (unter der Bedingung A)* durch

$$\boxed{P(B|A) \overset{\text{def}}{=} \frac{P(B\cap A)}{P(A)}}$$

Dadurch wird der Begriff (gegenüber der intuitiven Vorstellung) vielleicht wieder um einiges abstrakter (und auch leistungsfähiger), aber man sollte doch auch die einfache Vorstellung von *Wahrscheinlichkeiten in Spezialfällen* bei bedingten Wahrscheinlichkeiten bewahren: „$P(B|D)$ = Wahrscheinlichkeit des Ereignisses B (Brillenträger) in jenem *Spezialfall*, wenn man schon weiß, daß das Ereignis D (Dame) eintritt bzw. bereits eingetreten ist". Sollte sich bei zwei Ereignissen A,B herausstellen, daß $P(A|B) = P(A)$ ist, so ist die Wahrscheinlichkeit des Ereignisses A offenbar unabhängig davon, ob B eingetreten ist oder nicht: die Zusatzinformation, daß B eingetreten ist, bringt bzgl. der Wahrscheinlichkeit des Ereignisses A keine Veränderung. Es scheint daher auch anschaulich naheliegend, zu sagen bzw. zu definieren: Zwei Ereignisse A und B heißen *(stochastisch) unabhängig* voneinander, wenn $P(A|B) = P(A)$ gilt.

Eine wichtige „Rechenregel" im Zusammenhang mit bedingten Wahrscheinlichkeiten ist der „Satz von der totalen Wahrscheinlichkeit". Er dient der Berechnung von „totalen" Wahrscheinlichkeiten durch gewisse „Teilwahrscheinlichkeiten" (dies sind eben bedingte Wahrscheinlichkeiten). Das Ereignis B = „es wird ein Brillenträger gewählt" (als „totales" Ereignis oder „Gesamtereignis") hat – wie wir von oben schon wissen – eine Wahrscheinlichkeit von $P(B) = \frac{5}{14}$. Diese Wahrscheinlichkeit kann aber auch durch „Aufspalten der totalen Wahrscheinlichkeit in Teilwahrscheinlichkeiten, also durch Fallunterscheidung" berechnet werden, genau dies besagt – salopp, aber u.E. hilfreich formuliert – der Satz von der totalen Wahrscheinlichkeit.

Satz von der totalen Wahrscheinlichkeit: Ist X eine diskrete Zufallsvariable und A_n ($n = 1,2,\ldots,m$) eine *vollständige Zerlegung des Ereignisraumes* Ω [auch *vollständiges Ereignissystem* oder *vollständige Ereignisdisjunktion* genannt, d.h. ein Ereignissystem mit $\bigcup_{n=1}^{m} A_n = \Omega,\ A_n \neq \emptyset$ und $A_i \cap A_j$ für $i \neq j$], dann gilt

$$P(B) = P(B|A_1)\,P(A_1) + \ldots + P(B|A_m)\,P(A_m). \qquad (1)$$

Im Fall einer Aufteilung von Ω in zwei disjunkte Ereignisse A und $\neg A$ bedeutet dies $P(B) = P(B|A)\cdot P(A) + P(B|\neg A)\cdot P(\neg A)$. In unserem Beispiel ergibt sich dadurch

$$P(B) = P(B|D)P(D) + P(B|H)P(H) = \frac{2}{7}\cdot\frac{1}{2}+\frac{3}{7}\cdot\frac{1}{2} = \frac{5}{14}.$$

Bei diesem „Berechnen von Wahrscheinlichkeiten durch Fallunterscheidung" müssen die jeweiligen Teilwahrscheinlichkeiten (bedingte Wahrscheinlichkeiten in gewissen Sonderfällen, hier: D bzw. H) *gewichtet* werden nach der Wahrscheinlichkeit, mit der die jeweiligen Sonderfälle überhaupt eintreten.

So gesehen besagt der Satz von der totalen Wahrscheinlichkeit nichts anderes, als daß eine Gesamtwahrscheinlichkeit als **gewichtetes Mittel einzelner zugehöriger Teilwahrscheinlichkeiten** (Wahrscheinlichkeiten in Sonderfällen, Gewichtung je nach Wahrscheinlichkeit der Sonderfälle) berechnet werden kann.

Diese Sichtweise hat auch eine Entsprechung im deterministischen Bereich, bei *Anteilen, Prozentsätzen* bzw. *relativen Häufigkeiten*: Berechnen von Gesamt-*Anteilen, Prozentsätzen* bzw. *relativen Häufigkeiten* durch Gewichtung der entsprechenden (Teil-)Werte in den jeweiligen „Teilen" bzw. „Teilgruppen"; z.B. Prozentsätze:

$$25\% \text{ von } 80l + 50\% \text{ von } 120l = 40\% \text{ von } 200l.$$

Der Gesamtprozentsatz 40 kann als gewichtetes Mittel der Teilprozentsätze 25 und 50 aufgefaßt werden:

$$\underbrace{25\cdot\frac{80}{200}}_{10}+\underbrace{50\cdot\frac{120}{200}}_{30} = 40.$$

Bevor wir uns nun dem Begriff „Bedingter Erwartungswert" zuwenden, sei kurz auch die Definition des Erwartungswertes einer diskreten Zufallsvariable angegeben.

Definition: Es sei X eine diskrete Zufallsvariable mit $\{x_1, x_2, x_3, \ldots, x_k, \ldots\}$ als mögliche Werte. Unter dem Erwartungswert $E(X)$ der Zufallsvariable X versteht man die Summe

$$\boxed{E(X) \overset{\text{def}}{=} \sum_k x_k P(X = x_k),} \qquad (2)$$

falls diese (möglicherweise unendliche) Reihe absolut konvergiert. Andernfalls sagt man, die Zufallsvariable X hat keinen Erwartungswert.

Der *Erwartungswert* einer diskreten Zufallsvariable X ist also nichts anderes als ein *(gewichteter) Mittelwert* aller möglicher Ausprägungen x_k, gewichtet mit den jeweiligen Wahrscheinlichkeiten $P(X = x_k)$, deren Summe natürlich 1 ergibt.

Oft ist es auch günstig, jene Sichtweise zugrundezulegen, nach der eine Zufallsvariable eine (meßbare) Funktion vom Ereignisraum Ω nach R ist (X: $\Omega \to$ R), und statt von den Werten x_k von den (Funktions-) Werten $X(\omega_i)$zu sprechen. In dieser Sichtweise kann der Erwartungswert einer diskreten Zufallsvariable X auch folgendermaßen geschrieben bzw. definiert werden (mit $\Omega = \{\omega_1, ..., \omega_i, ...\}$), wobei die Äquivalenz zu obiger Definition wohl evident ist. Falls die (möglicherweise unendliche) Reihe absolut konvergiert, definiert man:

$$\boxed{E(X) \overset{\text{def}}{=} \sum_i X(\omega_i)P(\{\omega_i\}) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}).} \quad (3)$$

Beispiel: Sei X die Augenzahl beim Werfen eines Würfels. Der Ereignisraum Ω besteht dann aus den möglichen Versuchsergebnissen $\Omega = \{1,2,3,4,5,6\}$, wobei hier die einzelnen ω_i gleich den einzelnen x_i, den möglichen Werten der Zufallsvariable sind. Da die Wahrscheinlichkeit für jeden möglichen Versuchsausgang beim Würfeln $\frac{1}{6}$ beträgt, ist der Erwartungswert von X (bekanntlich) durch $E(X) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = 3.5$ gegeben.

1.2 „Bedingte Erwartungswerte" – Einführung und Definition für diskrete Zufallsvariable

Wenn in einem Stochastikkurs gewöhnliche Erwartungswerte und bedingte Wahrscheinlichkeiten thematisiert werden, dann ist es kein großer Schritt mehr zum Begriff des *bedingten Erwartungswertes*, der auch keine großen Verständnisschwierigkeiten machen sollte, zumindest dann nicht, wenn *Erwartungswerte* und *bedingte Wahrscheinlichkeiten* bereits verstanden wurden. Es bedarf lediglich eines Vereinens bzw. Vernetzens der gelernten Begriffe.

Beispiel: Die Zufallsvariable X stelle die Augenzahl beim Würfeln dar und A bedeute „es fällt eine gerade Zahl", dann wird man auch intuitiv [wenn man nur den Begriff *Erwartungswert*, nicht aber *bedingte Erwartungswerte* kennt] den *Erwartungswert von X, wenn man weiß, daß eine gerade Zahl gefallen ist* (diese Information kann z.B. von einem anderen Beobachter stammen) durch folgende Überlegung gewinnen: Durch die Information A = „es ist eine gerade Augenzahl gefallen" wird der ursprüngliche Ereignisraum $\Omega = \{1,2,3,4,5,6\}$ auf $\Omega \cap A = \{2,4,6\}$ eingeschränkt, es kommen nur mehr noch diese in Frage (analog zu den „Damen" bei den bedingten Wahrscheinlichkeiten). Da keiner dieser drei noch möglichen Ausgänge des Würfelexperimentes durch die gegebene Information bevorzugt wird, wird man wohl jedem Wahrscheinlichkeit 1/3 zuordnen. Man wird also den in Rede stehenden (bedingten) Erwartungswert (in diesem „Sonderfall", unter dieser „Bedingung") intuitiv durch gewöhnliche Erwartungswertbildung unter den vermöge der Bedingung weiterhin möglichen Werten gewinnen und ihn mit

$$„E(X|A)\text{"} = 2\cdot\frac{1}{3} + 4\cdot\frac{1}{3} + 6\cdot\frac{1}{3} = 4$$

angeben (die Anführungszeichen sollen andeuten, daß an dieser Stelle der Begriff $E(X|A)$ noch gar nicht definiert ist). Ganz analog wie bei bedingten Wahrscheinlichkeiten kann dieser intuitive Zugang („bedingter Erwartungswert = Erwartungswert in einem Spezial- bzw. Sonderfall", „Einschränkung des Ereignisraumes Ω und dann (gewichtete) Mittelwertbildung der verbleibenden Werte") wieder exaktifiziert werden und in einer Definition münden (siehe unten).

Salopp formuliert: Wenn X eine Zufallsvariable und A ein Ereignis (mit positiver Wahrscheinlichkeit) ist, dann ist der bedingte Erwartungswert $E(X|A)$ der Zufallsvariable X der Erwartungswert von X, wenn man weiß, daß das Ereignis A eingetreten ist. Man betrachtet also nur mehr solche Fälle (läßt nur mehr diese zu), die das Ereignis A realisieren, und bestimmt unter dieser *Bedingung* den Erwartungswert der Zufallsvariablen X. Es passiert also nichts weiter, als daß der Ereignisraum Ω auf $\Omega \cap A$ reduziert wird, d.h. auf jene Ereignisse, die A realisieren (dann „gewöhnliche" Erwartungswertbildung).

Definition: Es sei X eine diskrete Zufallsvariable mit $\{x_1, x_2, x_3, ..., x_k, ...\}$ als mögliche Werte und A ein Ereignis mit positiver Wahrscheinlichkeit. Der *bedingte Erwartungswert* bezüglich der Bedingung A wird dann definiert durch

$$\boxed{E(X|A) \overset{\text{def}}{=} \sum_k x_k P(X = x_k | A),} \qquad (4)$$

falls die (möglicherweise unendliche) Reihe absolut konvergiert. In der Definition sind also die Wahrscheinlichkeiten (im Vergleich zum gewöhnlichen Erwartungswert) durch bedingte Wahrscheinlichkeiten zu ersetzen.

Bemerkungen:

(1) Die oben auftretenden Wahrscheinlichkeiten mit dem Wert 1/3 sind also nichts anderes als die bedingten Wahrscheinlichkeiten

$P(2|A) = \frac{1}{3}$, $P(4|A) = \frac{1}{3}$, $P(6|A) = \frac{1}{3}$.

(2) In der Definition des bedingten Erwartungswertes wird zwar über alle möglichen Ausprägungen x_k der Zufallsvariable X summiert, also in unserem Beispiel auch über die Werte 1,3,5, aber die zugehörigen bedingten Wahrscheinlichkeiten haben alle den Wert 0: $P(1|A) = P(3|A) = P(5|A) = 0$. Der obige intuitive Zugang (Einschränkung von Ω, dann gewöhnliche Erwartungswertbildung) ist also mit der hier gegebenen Definition (4) „verträglich". Dies ist zwar einfach einzusehen, darf aber nicht von vornherein als gegeben angenommen werden; es ist eine zu beweisende Tatsache (siehe unten formulierten Satz). Wir bedienen uns zur Erläuterung dessen jener Schreibweise, bei der die Zufallsvariable X als Funktion von Ω nach R gesehen wird:

$$\boxed{E(X|A) \overset{\text{def}}{=} \sum_{\omega \in \Omega} X(\omega)P(\omega|A),} \qquad (5)$$

Die Wahrscheinlichkeit jedes Versuchsausganges $P(i)$ (für jedes $i \in \{1,2,3,4,5,6\}$) hatte beim ursprünglichen Erwartungswert jeweils den Wert $\frac{1}{6}$.

Bei der Berechnung von $E(X|A)$ haben wir nur mehr die geraden Zahlen 2,4,6 zugelassen und die „zugehörigen Wahrscheinlichkeiten haben sich von jeweils $\frac{1}{6}$ auf jeweils $\frac{1}{3}$ erhöht". Diese Vergrößerung der „Gewichte" (bei der Mittelwertbildung) um den Faktor 2 ist zwar schon a priori recht einsichtig, da in Summe die Gewichte natürlich wieder 1 sein müssen, sich aber die Anzahl der Gewichte halbiert hat, bedarf aber eben doch einer genaueren Begründung:

Satz: Für den bedingten Erwartungswert $E(X|A)$ einer diskreten Zufallsvariable unter der Bedingung A gilt:

$$E(X|A) = \sum_{\omega \in A} X(\omega)\frac{P(\{\omega\})}{P(A)} = \frac{1}{P(A)}\sum_{\omega \in A} X(\omega)P(\{\omega\}).$$

Dieser Satz besagt also genau das oben Beschriebene: die intuitive Sichtweise ist „verträglich" mit der Definition: (1) Einschränkung in der Summe auf $\omega \in A$ und (2) Erhöhung der zugehörigen Wahrscheinlichkeiten um den Faktor $\frac{1}{P(A)} > 1$ [d.h. wenn A ein Drittel von Ω ausmacht, so werden die Gewichte verdreifacht; wenn – wie oben – A die Hälfte von Ω ausmacht, so müssen die zugehörigen Wahrscheinlichkeiten (= Gewichte) verdoppelt werden].

Beweis: Nach Definition für bedingte Wahrscheinlichkeiten gilt

$$P(\omega|A) \overset{\text{def}}{=} \frac{P(\{\omega\}\cap A)}{P(A)} = \begin{cases} \frac{P(\{\omega\})}{P(A)}, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A \end{cases}.$$

Laut Definition des bedingten Erwartungswertes in der Sichtweise von (5) ist

$$E(X|A) \overset{\text{def}}{=} \sum_{\omega \in \Omega} X(\omega)P(\omega|A),$$

woraus sich mit obiger Darstellung der bedingten Wahrscheinlichkeiten $P(\omega\,|\,A)$ die Behauptung des Satzes unmittelbar ergibt.

Durch die Kombination *bedingte Wahrscheinlichkeit* und *Erwartungswert* zum *bedingten Erwartungswert* wird in einem gewissen Sinn nichts geschaffen, was neu, überraschend oder viel schwieriger zu verstehen wäre – es wird lediglich (wie bei bedingten Wahrscheinlichkeiten) der Ereignisraum eingeschränkt. Unter diesem Aspekt ist auch völlig klar, daß bedingte Erwartungswerte alle bekannten Eigenschaften von gewöhnlichen Erwartungswerten haben, z.B. die Linearität: $E(aX + Y\,|\,A) = a\,E(X\,|\,A) + E(Y\,|\,A)$.

1.3 Der Satz vom totalen Erwartungswert

In Analogie zum **Satz von der totalen Wahrscheinlichkeit** (1) gilt auch der

Satz vom totalen Erwartungswert: Ist X eine diskrete Zufallsvariable und A_n ($n = 1, 2, ..., m$) eine *vollständige Zerlegung des Ereignisraumes* Ω, so gilt

$$(6) \qquad \boxed{\begin{aligned} E(X) &= E(X|A_1)P(A_1) + ... + E(X|A_m)\,P(A_m) \\ &= \sum_{n=1}^{m} E(X|A_n)P(A_n). \end{aligned}}$$

Beweis: Aufgrund des Satzes von der totalen Wahrscheinlichkeit gilt für das Ereignis $X = x_k$:

$$P(X = x_k) = \sum_n P(X = x_k\,|\,A_n)P(A_n)$$

und damit

$$E(X) = \sum_k x_k P(X = x_k) =$$

$$\sum_k x_k \sum_n P(X = x_k\,|\,A_n)P(A_n) =$$

$$\sum_n P(A_n)\underbrace{\sum_k x_k P(X = x_k\,|\,A_n)}_{E(X|A_n)} = \sum_n E(X\,|\,A_n)P(A_n),$$

was ja zu zeigen war. In der Schule wäre es wahrscheinlich besser, zuerst einen entsprechenden Satz mit einer Teilung von Ω in zwei disjunkte Ereignisse

A_1 und $A_2 = \neg A_1$ zu behandeln (nicht nur formal, sondern vor allem auch inhaltlich, um durch einfache Beispiele Verständnis und richtige Vorstellungen zu generieren).

Beispiel: Sei wie oben X die Augenzahl beim Werfen eines Würfels, A_1 das Ereignis „es erscheint eine gerade Augenzahl" und A_2 das Ereignis „es erscheint eine ungerade Augenzahl". Dann ist der bedingte Erwartungswert $E(X|A_1) = 4$ (siehe oben) und $E(X|A_2) = 1\cdot\frac{1}{3} + 3\cdot\frac{1}{3} + 5\cdot\frac{1}{3} = 3$. Nach dem Satz vom totalen Erwartungswert erhalten wir

$$E(X) = E(X|A)\cdot P(A_1) + E(X|A_2)\cdot P(A_2) =$$
$$4\cdot\frac{1}{2} + 3\cdot\frac{1}{2} = 3.5$$

Der Satz vom totalen Erwartungswert besagt nichts anderes, als daß der Erwartungswert $E(X)$ (auch) durch „Berechnen des Erwartungswertes durch Fallunterscheidung" bzw. „Bestimmen von Erwartungswerten in Sonder- bzw. Spezialfällen (Teilerwartungswerte)" gewonnen werden kann. Der Gesamterwartungswert ist, so gesehen, ein gewichtetes Mittel aus einzelnen Erwartungswerten in Sonderfällen (bedingte Erwartungswerte), wobei diese Sonderfälle eben durch die Ereignisse A_i angezeigt werden, und diese „Teilerwartungswerte" nach der Auftrittswahrscheinlichkeit der Ereignisse A_i gewichtet werden müssen.

Auch hier gibt es eine Analogie zum Deterministischen: Berechnen von Mittelwerten durch Gewichtung von „Teilmittelwerten": Der (Gesamt-) Mittelwert m der Körpergrößen von N Personen kann auch durch Berechnung der Teilmittelwerte m_1 und m_2 in zwei Teilgruppen zu N_1 bzw. N_2 Personen geschehen, wobei diese Teilmittelwerte mit der jeweiligen Gruppengröße zu gewichten sind:

$$m = m_1\cdot\frac{N_1}{N} + m_2\cdot\frac{N_2}{N}\ ,\ N_1 + N_2 = N.$$

Analog bei einer Aufteilung in k Gruppen:

$$m = m_1\cdot\frac{N_1}{N} + \ldots + m_k\cdot\frac{N_k}{N}\ ,\ N_1 + \ldots + N_k = N.$$

1.4 Anwendung: Einfachere Herleitung des Erwartungswertes einer geometrisch verteilten Zufallsvariable: $E(X) = 1/p$

Eine besonders fruchtbare Anwendung bedingter Erwartungswerte ist die Berechnung des Erwartungswertes einer *geometrisch* verteilten Zufallsvariable, die sonst doch einigen „Aufwand" erfordert (unendliche Reihen, gliedweises Differenzieren etc.). Sei also X geometrisch verteilt mit dem Parameter p, d.h. X sei die Anzahl der notwendigen Versuchswiederholungen eines Zufallsexperiments bis zum *ersten* Eintreten eines bestimmten Ereignisses A, das bei jeder Versuchswiederholung mit konstanter Wahrscheinlichkeit p eintritt (Unabhängigkeit der einzelnen Versuche sei vorausgesetzt – „BERNOULLI-Experiment"). Die beiden einfachsten und am öftesten genannten Beispiele dazu sind wohl: „Werfen einer Münze, bis zum ersten Mal *Kopf* fällt" oder „Würfeln, bis zum ersten Mal eine *Sechs* fällt".

Nun gilt nach obigem Satz vom totalen Erwartungswert (mit A sei hier gemeint: „beim ersten Versuch tritt A ein")

$$\begin{aligned} E(X) &= E(X|A)P(A) + E(X|\neg A)P(\neg A) \\ &= 1\cdot p + (1 + E(X))(1-p). \end{aligned} \tag{7}$$

Die hier eingegangene Identität $E(X|\neg A) = 1 + E(X)$ ist leicht zu begründen: Wenn beim ersten Versuch nicht A (also $\neg A$) eintritt, ist bereits *ein Fehl*versuch passiert (daher „1 +...") und das Spiel beginnt von neuem (daher „... + $E(X)$"). Aus (7) folgt sofort $pE(X) = 1$ bzw. das endgültige Resultat

$$\boxed{E(X) = \frac{1}{p}.} \tag{8}$$

Im Durchschnitt (auf lange Sicht) muß man demnach zweimal eine Münze werfen, um zum ersten Mal *Kopf*, bzw. sechsmal einen Würfel, um erstmalig eine Sechs zu erhalten.

Bemerkung: Die gegebene Begründung für $E(X|\neg A) = 1 + E(X)$ scheint uns zwar ohnehin ausreichend zu sein, sie läßt sich aber auch „formal exakt" durch Einsetzen in die Definition eines Erwartungswertes (2) bzw. bedingten Erwartungswertes (4) bewerkstelligen.

$$E(X) \overset{\text{def}}{=} \sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}p =$$
$$1\cdot p + 2\cdot(1-p)p + 3\cdot(1-p)^2p + 4\cdot(1-p)^3p + \ldots$$
$$E(X|\neg A) \overset{\text{def}}{=} \sum_{k=2}^{\infty} k\cdot(1-p)^{k-2}p =$$
$$2\cdot p + 3\cdot(1-p)p + 4\cdot(1-p)^2p + 5\cdot(1-p)^3p + \ldots$$

Daraus erkennt man
$E(X|\neg A) = E(X+1) = E(X) + 1.$

2 Weitere Beispiele zum bedingten Erwartungswert

In diesem Abschnitt wird eine Reihe von Beispielen vorgestellt, bei denen die Berechnung des totalen Erwartungswertes im Vordergrund steht. Die dabei auftretenden bedingten Erwartungswerte können entweder direkt und intuitiv berechnet werden (dann mit deren Hilfe auch der gefragte Erwartungswert), oder durch den Satz vom totalen Erwartungswert entsteht eine einfache Gleichung für den unbekannten Erwartungswert. Die Existenz der jeweiligen Erwartungswerte wird in keinem Fall thematisiert, sondern als gegeben vorausgesetzt.

Beispiel 2.1 *Bei einer Schießbude (mit relativ wertvollen Gewinnen, daher sind die Kosten vergleichsweise etwas höher) sind die Kosten für die einzelnen Schießversuche an einer Tafel abzulesen:*

1. Versuch:	*100,– DM*
2. Versuch:	*80,– DM*
3. Versuch:	*60,– DM*
4. Versuch:	*40,– DM*
ab 5. Versuch:	*20,– DM*

Jemand trifft unabhängig von den einzelnen Versuchen jeweils mit Wahrscheinlichkeit p (die Wahrscheinlichkeit eines Fehlschusses ist daher $q \overset{\text{def}}{=} 1-p$) und schießt solange, bis er einen Treffer erzielt. Was werden die mittleren Kosten einer solchen Serie sein?

Wenn er schon beim ersten Mal trifft, so betragen die Kosten 100 DM (die Wahrscheinlichkeit dafür ist p), wenn er erst beim zweiten Mal trifft, so betragen die Kosten 180 DM (dies ist mit Wahrscheinlichkeit qp der Fall), wenn er erst beim dritten Versuch trifft (Wahrscheinlichkeit $= q^2p$), so belaufen sich die Kosten auf 240 DM und wenn er erst beim vierten Versuch trifft (Wahrscheinlichkeit $= q^3p$), so hat er 280 DM zu bezahlen. Wenn wir wissen, daß der Schütze bereits vier Fehlversuche hinter sich hat, dann hat er schon 100 + 80 + 60 + 40 = 280,– DM bezahlt und der Erwartungswert der Anzahl der noch nötigen Versuche beträgt (geometrische Verteilung) $\frac{1}{p}$. Daher beträgt der Erwartungswert der Kosten dann $280 + \frac{20}{p}$ DM. Die (zugehörige) Wahrscheinlichkeit, *daß* er vier Fehlversuche hat, ist q^4. Eine übersichtliche Zusammenstellung ist in Tab. 2 gegeben.

Zerlegung von Ω in $A_1,\dots,A_5$	E(Kosten|*)	Wahrscheinlichkeit
$A_1 \hat{=}$ Treffer beim 1. Mal:	100,–	p
$A_2 \hat{=}$ 1. Treffer beim 2. Mal:	180,–	qp
$A_3 \hat{=}$ 1. Treffer beim 3. Mal:	240,–	q^2p
$A_4 \hat{=}$ 1. Treffer beim 4. Mal:	280,–	q^3p
$A_5 \hat{=}$ 1. Treffer erst ab 5. Mal:	$280 + \frac{20}{p}$	q^4

Tab. 2: Kosten in Abhängigkeit vom 1. Treffer

Durch Aufspalten des Erwartungswertes in fünf bedingte Erwartungswerte kommt man zum Ergebnis:

$$\boxed{E(\text{Kosten}) = 100p + 180qp + 240q^2p + 280q^3p + \left(280 + \frac{20}{p}\right)q^4 \text{ DM.}} \tag{9}$$

Bemerkung: Wenn p nicht die Trefferwahrscheinlichkeit, sondern die Wahrscheinlichkeit für einen Fehlschuß ist, so werden durch (9) die mittleren Kosten einer Serie bis zum ersten Fehlschuß dargestellt.

Beispiel 2.2 *Der Gewinner eines Preisausschreibens steht vor einem Kästchen mit vier Schubladen. Eine Schublade enthält 3000 DM, eine zweite 2000 DM, eine dritte 1000 DM und die vierte ist die „Nietenlade", sie enthält keinen Gewinn. Der Gewinner darf nun eine Lade ziehen und den darin enthaltenen Betrag als Gewinn verbuchen. Nun wird erstens die geleerte Lade wieder mit demselben Geldbetrag gefüllt und zweitens werden die Laden durch den Quizmaster (oder durch einen Zufallsgenerator) neu im Kästchen verteilt, so daß der Gewinner dies nicht sieht und er ein zweites Mal das Glück herausfordern kann („2. Runde"). Dies wird solange wiederholt, bis der Gewinner die „Nietenlade" erwischt. Wieviel muß die Lotteriegesellschaft dem Gewinner eines solchen Preisausschreibens im Mittel bezahlen?*

Da der Gewinner bei allen Ladenöffnungen keinerlei Information hat, wählt er bei jedem Versuch eine der vier Laden mit Wahrscheinlichkeit $\frac{1}{4}$. Wenn er beim ersten Versuch die Lade mit 3000 DM erwischt, so hat ihm diese Runde eben 3000 DM eingebracht und er steht vor der gleichen Situation, vor der er ganz zu Beginn stand. Wenn E der (unbekannte) Erwartungswert des Gewinnes dieses Spiels ist, so ist dann sein nunmehriger Erwartungswert wohl durch 3000 + E gegeben (die zugehörige Wahrscheinlichkeit beträgt $\frac{1}{4}$). Analoge Überlegungen für die anderen möglichen Laden bei seiner ersten Ziehung führen zur Gleichung

$$\boxed{E = \tfrac{1}{4}(3000 + E) + \tfrac{1}{4}(2000 + E) + \tfrac{1}{4}(1000 + E) + \tfrac{1}{4}\cdot 0,}$$

woraus wir unmittelbar $E = 6000$ DM erhalten. Dies ist auch durch folgende Überlegung plausibel: Auf lange Sicht wird die vierte Ladenöffnung erstmals die „Nietenlade" bringen (Erwartungswert der geometrischen Verteilung!) und vorher wird jede Lade einmal drangekommen sein: 3000 + 2000 + 1000 = 6000.

Beispiel 2.3 *Wie groß ist beim Würfeln die erwartete Augensumme, bevor zum ersten Mal eine „6" fällt?*

Dies ist ein völlig analoges Beispiel zu 2.2. Diesmal wollen wir es formaler lösen. Sei AS die Augensumme, bevor die erste „6" fällt, 1 sei das Ereignis „beim ersten Wurf fällt eine 1" (analog für 2,3,4,5,6). Nach dem Satz vom totalen Erwartungswert erhalten wir:

$$E(AS) = E(AS|\boxed{1})\cdot P(\boxed{1}) + E(AS|\boxed{2})\cdot P(\boxed{2}) + E(AS|\boxed{3})\cdot P(\boxed{3}) + E(AS|\boxed{4})\cdot P(\boxed{4}) + E(AS|\boxed{5})\cdot P(\boxed{5}) + E(AS|\boxed{6})\cdot P(\boxed{6}).$$

Nun ist klar, daß $P(\boxed{i}) = \frac{1}{6}$ für $i = 1,\ldots,6$ und $E(AS|\boxed{i}) = i + E(AS)$ für $i = 1,\ldots,5$ bzw. $E(AS|\boxed{6}) = 0$ gilt.

Damit erhalten wir mit $E \overset{\text{def}}{=} E(AS)$

$$\boxed{E = \frac{1}{6}\,[(1+E)+(2+E)+(3+E)+(4+E)+(5+E)],}$$

und daraus schließlich $E = 1 + 2 + 3 + 4 + 5 = 15$.

Beispiel 2.4 *Was wäre der faire Einsatz für folgendes Spiel: In einer Ziehungstrommel befinden sich 20 Kugeln: 2 rote, 3 grüne, 4 blaue und 11 weiße. Vor jeder Ziehung werden die Kugeln mittels eines Gebläses durchmischt; die gezogene Kugel wird nach jeder Ziehung wieder „zurückgelegt", so daß vor jeder Ziehung dieselbe Ausgangslage herrscht. Eine rote Kugel bringt einen Gewinn von 1000 DM, eine grüne eine Gewinn von 500 DM, eine blaue Kugel bringt 100 DM und eine weiße Kugel bringt keinen Gewinn. Der Spieler darf solange die Mischmaschine für eine neuerliche Ziehung starten, bis er eine weiße Kugel erwischt und damit keinen Gewinn erzielt.*

Wenn man unter dem fairen Einsatz den Erwartungswert E des Gewinnes versteht, so erhält man mit Hilfe des Satzes vom totalen Erwartungswert für E die Beziehung

$$\boxed{E = \frac{2}{20}(1000+E) + \frac{3}{20}(500+E) + \frac{4}{20}(100+E) + \frac{11}{20}\cdot 0,}$$

woraus sich $E = \frac{3900}{11} \approx 354.55$ DM ergibt.

Beispiel 2.5 *Ein Arbeiter bedient in einer Fabrikanlage n gleichartige Maschinen, die geradlinig und äquidistant angeordnet sind (siehe Abb. 1). Der Abstand je zweier benachbarter Maschinen betrage a, die Gesamtlänge l dieser „Maschinenstraße" daher l = (n − 1)a. Die Maschinen arbeiten an sich völlig selbständig, nur in einem Störfall, der durch ein Aufleuchten einer Lampe angezeigt wird, muß der Arbeiter zur jeweiligen Maschine gehen und nach dem Rechten sehen. Welchen Weg muß der Arbeiter beim Gang zu einer aufleuchtenden Maschine „im Durchschnitt" zurücklegen (Erwartungswert der Weglänge zwischen den Maschinen)?*

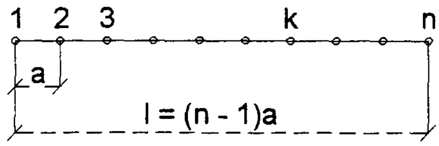

Abb. 1: n Maschinen im Abstand a

Wir numerieren die Maschinen von links nach rechts von 1 bis n; der Arbeiter stehe gerade bei der Maschine mit Nummer k (wir bezeichnen dieses Ereignis mit M_k). Da der Arbeiter keine Vorliebe für eine bestimmte Maschine hat, ist die Wahrscheinlichkeit, daß er sich gerade bei der k-ten Maschine befindet, gleich $\frac{1}{n}$, also $P(M_k) = \frac{1}{n}$ für $1 \le k \le n$.

Da alle Maschinen nach Voraussetzung von der gleichen Art sind, ist die Wahrscheinlichkeit dafür, daß die nächste Maschine, welche die Aufmerksamkeit des Arbeiters erfordert, die Nummer i trägt, in jedem Fall auch gleich $\frac{1}{n}$ $(1 \le i \le n)$ (auch für $i = k$, dann hätte er gar keinen Weg zurückzulegen). Wenn sich also der Arbeiter bei der Maschine k befindet und der nächste Störanfall bei der Maschine i stattfindet, so ist die zurückzulegende Wegstrecke $W|M_k$ (Weglänge, wenn man weiß, daß sich der Arbeiter bei der Maschine k befindet) gegeben durch

$$W|M_k = \begin{cases}(k-i)a & \text{für } i \le k,\\ (i-k)a & \text{für } i > k.\end{cases}$$

$W|M_k$ ist eine Zufallsvariable, die (bei festem k) von i abhängig ist. Für ihren Erwartungswert erhalten wir nach Definition (der konstante Wert $\frac{1}{n}$ wurde bereits herausgehoben: Wahrscheinlichkeit, daß die nächste Störung bei der Maschine i stattfindet):

$$\begin{aligned}E(W|M_k) &= \frac{1}{n}\left(\sum_{i=1}^{k}(k-i)a + \sum_{i=k+1}^{n}(i-k)a\right)\\ &= \frac{a}{n}\left(\frac{k(k-1)}{2} + \frac{(n-k)(n-k+1)}{2}\right)\\ &= \frac{a}{2n}\left[2k^2 - 2(n+1)k + n(n+1)\right].\end{aligned}$$

Mit $P(M_k) = \frac{1}{n}$ und dem Satz vom totalen Erwartungswert [hier: $E(W) = \sum_{k=1}^{n} E(W|M_k)\cdot P(M_k)$] erhalten wir für den interessierenden Erwartungswert:

$$E(W) = \sum_{k=1}^{n} \frac{a}{2n^2}\left[2k^2 - 2(n+1)k + n(n+1)\right].$$

Bekanntlich ist $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ (oben schon benutzt)

und $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, wodurch sich nach wenigen Umformungen

$$\boxed{E(W) = \frac{a(n^2-1)}{3n} = \frac{l}{3}\left(1+\frac{1}{n}\right)}$$

ergibt (dabei bezeichnet $l = (n-1)a$ den Abstand der beiden äußersten Maschinen). Bei gleichbleibendem l und wachsendem n konvergiert dieser Erwartungswert also von $l\cdot\frac{1}{2}$ (bei $n = 2$) streng monoton fallend nach $l\cdot\frac{1}{3}$ (z.B. $l\cdot\frac{5}{12}$ bei $n = 4$, oder $l\cdot\frac{11}{30}$ bei $n = 10$).

Beispiel 2.6 *Ein schwarzer und ein roter Würfel werden geworfen. Wie groß ist der durchschnittliche Unterschied der beiden Augenzahlen?*

Sei Δ die Zufallsvariable, die den Unterschied der beiden gewürfelten Augenzahlen beschreibt. Ihre möglichen Ausprägungen sind natürlich $\{0,1,2,3,4,5\}$; der größte Unterschied 5 ergibt sich bei (1,6) bzw. (6,1), der kleinste Unterschied 0 ergibt sich bei gleichen Augenzahlen, also bei (1,1), (2,2), (3,3), (4,4), (5,5), (6,6). Wir interessieren uns für den Erwartungswert der Zufallsvariable Δ.

Die vollständige Ereignisdisjunktion nehmen wir vor nach der vom schwarzen Würfel gezeigten Augenzahl: diese kann 1,2,3,4,5,6 sein. Sei also S_k das Ereignis, daß der schwarze Würfel k zeigt, $k \in (1,\ldots,6)$.

Nach dem Satz vom totalen Erwartungswert erhalten wir zunächst

$$E(\Delta) = E(\Delta|S_1)P(S_1) + \ldots + E(\Delta|S_6)P(S_6)$$

und mit $P(S_k) = \frac{1}{6}$ für $k \in \{1,\ldots,6\}$

$$E(\Delta) = \frac{1}{6}[E(\Delta|S_1)+\ldots+E(\Delta|S_6]] = \frac{1}{6}\sum_{k=1}^{6} E(\Delta|S_k).$$

Wir haben also zunächst die einzelnen bedingten Erwartungswerte $E(\Delta|S_k)$ zu berechnen, d.h. die Erwartungswerte von Δ in den einzelnen („Sonder-") Fällen, daß der schwarze Würfel k zeigt. Wenn der schwarze Würfel k (fest) zeigt, so hängt Δ natürlich nur mehr vom Ergebnis des roten Würfels ab. Je nachdem, ob der rote Würfel 1, 2, 3, 4, 5, 6 zeigt (jeweils mit Wahrscheinlichkeit $\frac{1}{6}$), ergeben sich die Werte von Δ und damit die (bedingten) Erwartungswerte in diesen einzelnen Fällen:

$$E(\Delta|S_1) = \frac{1}{6}\ (0+1+2+3+4+5) = \frac{15}{6}$$

$$E(\Delta|S_2) = \frac{1}{6}\ (1+0+1+2+3+4) = \frac{11}{6}$$

$$E(\Delta|S_3) = \frac{1}{6}\ (2+1+0+1+2+3) = \frac{9}{6}$$

$$E(\Delta|S_4) = \frac{1}{6}\ (3+2+1+0+1+2) = \frac{9}{6}$$

$$E(\Delta|S_5) = \frac{1}{6}\ (4+3+2+1+0+1\ ) = \frac{11}{6}$$

$$E(\Delta|S_6) = \frac{1}{6}\ (5+4+3+2+1+0) = \frac{15}{6}$$

Dadurch erhalten wir für den interessierenden Erwartungswert $E(\Delta)$

$$E(\Delta) = \frac{1}{6}\left(\frac{15}{6}+\frac{11}{6}+\frac{9}{6}+\frac{9}{6}+\frac{11}{6}+\frac{15}{6}\right) =$$

$$\frac{70}{36} = \frac{35}{18} = 1\frac{17}{18} \approx 1.94\,.$$

Der Erwartungswert des Unterschiedes der Augenzahlen beim Werfen zweier Würfel beträgt also knapp unter 2.

Literatur

Humenberger, H. (1998a): Ein überraschendes Phänomen bei Münzwurfserien und bedingte Erwartungswerte. Preprint Universität für Bodenkultur, Wien.

Humenberger, H. (1998b): Was haben Kopf-Adler-Muster in Münzwurfserien („binären Sequenzen") und FIBONACCI-Folgen miteinander zu tun? Preprint Universität für Bodenkultur, Wien.

Humenberger, H. (1998c): Der Erwartungswert der Augensumme beim „Würfeln mit Streichresultaten". Preprint Universität für Bodenkultur, Wien.

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

Anschrift des Verfassers:
Hans Humenberger,
Institut für Mathematik und Angewandte Statistik,
Universität für Bodenkultur,
Gregor Mendel-Straße 33,
A-1180 Wien.
E-mail: hans@edv1.boku.ac.at